

Syllabus

WIRELESS & MOBILE COMPUTING (VII-SEM.)

UNIT - 1 : REVIEW OF TRADITIONAL NETWORKS

Review of LAN, MAN, WAN, Intranet, Internet, and interconnectivity devices : bridges, routers, etc. Review of TCP/IP Protocol Architecture : ARP/RARP, IP addressing, IP datagram format and its delivery, routing table format, ICMP messages, subnetting, supernetting and CIDR, DNS. NAT : Private addressing and NAT, SNAT, DNAT, NAT and firewalls, VLANs : Concepts, comparison with real LANS, type of VLAN, tagging, IPV6 : address structure, address space and header.

UNIT - 2 : STUDY OF TRADITIONAL ROUTING AND TRANSPORT

Routing Protocols : BGP - Concept of hidden network and autonomous system, an exterior gateway protocol, different messages of BGP. Interior Gateway Protocol : RIP, OSPF. Multiplexing and Ports, TCP : Segment format, sockets, synchronization, three way hand shaking, variable window size and flow control, timeout and retransmission algorithms, connection control, silly window syndrome. Example of TCP : Tahoe, Reno, Sack, etc. UDP : Message Encapsulation, Format and Pseudo header.

UNIT - 3 : WIRELESS LAN

Transmission medium for WLANs, MAC problems, hidden and exposed terminals, near and far terminals, infrastructure and Ad hoc networks, IEEE 802.11-system arch, protocol arch, physical layer, concept of spread spectrum, MAC and its management, power management, security. Mobile IP : unsuitability of traditional IP, goals, terminology, agent advertisement and discovery, registration, tunneling techniques, Ad hoc network routing : Ad hoc network routing v/s traditional IP routing, types of routing protocols, Examples : AODV, DSDV, DSR, ZRP, etc.

UNIT - 4 : MOBILE TRANSPORT LAYER

Unsuitability of traditional TCP, I-TCP, S-TCP, M-TCP. Wireless Cellular Networks : Cellular system, cellular networks v/s WLAN, GSM – Services, system architecture, Localization and calling, handover and roaming.

UNIT - 5 : MOBILE DEVICE OPERATING SYSTEMS

Special constraints & requirements, commercial mobile operating systems. Software Development Kit : iOS, Android, etc. M-Commerce : Structure, Pros & Cons, Mobile payment system, security issues.

Price : Rs. 120.00 (Rs. One Hundred Twenty Only)
Edition : 2020

UNIT

1

REVIEW OF TRADITIONAL NETWORKS

REVIEW OF LAN, MAN, WAN, INTRANET, INTERNET, AND INTERCONNECTIVITY DEVICES – BRIDGES, ROUTERS, ETC.

Q.1. Write short notes on the following –

- (i) *Local area network (LAN)*
- (ii) *Metropolitan area network (MAN)*
- (iii) *Wide area network (WAN).*

(R.G.P.V., Dec. 2009)

Ans. (i) Local Area Network (LAN) – A Local Area Network (LAN) is usually privately owned and links the devices in a single office, building, or campus. Depending on the requirements of an organization and the type of technology used, a LAN can be as simple as two PCs and a printer in someone's home office, or it can extend throughout a company and include voice, sound, and video peripherals. LANs are differentiated from other types of networks by three characteristics –

- (a) Their size
- (b) Their transmission technology
- (c) Their topology.

Size of the LANs is restricted. It means that the worst-case transmission time is bounded and known in advance. By knowing this bound, it becomes possible to use certain types of designs which would not otherwise be possible. It also simplifies network management.

LANs often use a transmission technology which consists of a single cable to which all the machines are attached, like the telephone company party lines once used in rural areas. Traditional LANs run at speeds of 10 to 100 Mbps, have low delay (tens of microseconds), and make very few errors.

To broadcast LANs many topologies are possible. Two of them are shown in fig. 1.1. In a bus topology, at any moment one machine is the master and is permitted to transmit. All other machines are needed to refrain from sending. An arbitration mechanism is required to resolve conflicts when two or more machines want to transmit simultaneously. Arbitration mechanism can be centralized or distributed.

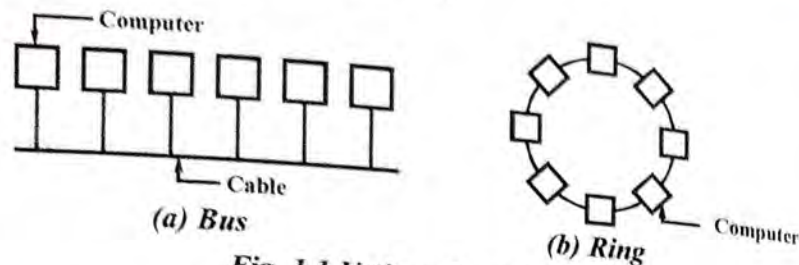


Fig. 1.1 Various Topologies

Ring is the second type of broadcast system. In a ring, each bit propagates around on its own, not waiting for the rest of the packet to which it belongs. Each bit circumnavigates the whole ring in the time it takes to transmit a few bits, often before the complete packet has even been transmitted.

Other type of LAN is built using point-to-point lines. Individual lines connect a specific machine with another specific machine. This type of LAN is really a miniature wide area network.

A local area network is generally a private owned communications network within a confined geographical area. The range is usually within a few kilometers. (less than 10 km) – may be one office, one building or a group of buildings located closely such as a college campus. The transmission channels generally use coaxial or fibre optic cables and special interface units rather than telephone lines and modems.

(ii) Metropolitan Area Network (MAN) – A Metropolitan Area Network (MAN) is basically a bigger version of a LAN and normally uses similar technology. It is designed to extend over an entire city. It may be a single network such as a cable television network or it may be a means of connecting a number of LANs into a larger network so that resources may be shared LAN-to-LAN as well as device-to-device. A key aspect of a MAN is that there is a broadcast medium to which all the computers are attached, due to which the design is greatly simplified as compared to there kind of networks.

A MAN can support both data and voice, and might even be related to the local cable television distribution network. It might cover a group of nearby offices or a city and might be either private or public. A MAN just has one or two cables and does not contain switching elements.

Main reason for even distinguishing MANs as a special category is that a standard has been adopted for them, and this standard is now being implemented. It is known as DQDB (Distributed Queue Dual Bus) or in number as 802.6. DQDB has two unidirectional buses (cables) to which all the computers are connected, as shown in fig. 1.2. Each bus has a head-end, a device which initiates transmission activity. Traffic that is destined for a computer to the right of the sender uses the upper bus and traffic to the left uses the lower one.

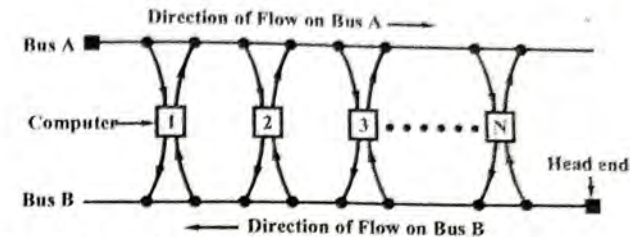


Fig. 1.2 Architecture of DQDB Metropolitan Area Network

(iii) Wide Area Networks – Wide Area Network (WAN) provides long-distance transmission of data, voice, image and video information over large geographical areas that may comprise a country, a continent, or even the whole world. It has a collection of machines intended for running user programs. These machines are called **hosts**. Hosts are connected by a **communication subnet** or just **subnet**. The subnet carries messages from host-to-host. By separating the pure communication aspects of the network (subnet) from the application aspects (hosts), the complete network design is greatly simplified.

There are two distinct components of the subnet is most wide area networks which are transmission lines and switching elements. Transmission lines move bits between machines. Switching elements are specialized computers which are used to connect two or more transmission lines.

In most WANs, the network has numerous cables or telephone lines, each one connecting a pair of routers. When two routers which do not share a cable nevertheless wish to communicate, they must perform this indirectly, via other routers. When a packet is transmitted from one router to another via one or more intermediate routers, the packet is received at each

intermediate router in its entirety, stored there until the needed output line is free, and then forwarded. A subnet using this principle is known as a **point-to-point, store-and-forward**, or **packet-switched** subnet.

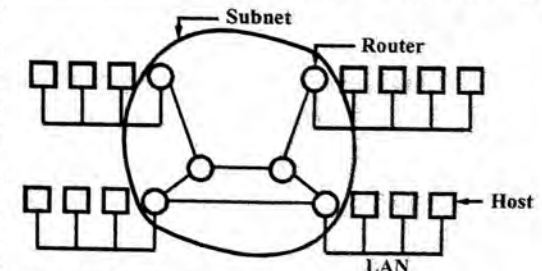


Fig. 1.3 Relation between Hosts and Subnet

A WAN can be a public system such as the Public Switched Telephone Network (PSTN) or one of the various packet switched services provided by the public telecommunication authorities WAN can also use most other types of circuits including satellite networks, ISDN, ATM, Value Added Networks. The main distinguishing feature between a LAN and a WAN is that the LAN is under the complete control of the owner.

Q.2. Write the range of LAN, WAN and MAN.

Ans. The range of LAN, WAN, and MAN –

LAN – LANs typically operate at bit rates from 10 Mbps to 100 Mbps within a relatively small geographical area like building.

WAN – WANs typically operate at speeds of 1.5 Mbps to 2.4 Gbps and cover a distance of 100 to 1000 miles.

MAN – MANs typically operate at bit rates from 1.5 Mbps to 10 Mbps and range from five miles to a few hundred miles in length.

Q.3. What is intranet? Give its advantages and disadvantages.

Ans. Intranet is the system in which multiple PCs are connected to each other. PCs in intranet are not available to the world outside the intranet. Usually, each organization has its own intranet network and members/employees of that organization can access the computers in their intranet.

Each computer in intranet is also identified by an IP address which is unique among the computers in that intranet.

Advantages of Intranet – Advantages of intranet are as follows –

- (i) Fast, easy, low-cost to implement
- (ii) Based on open standards
- (iii) Allows connectivity with other systems
- (iv) Access to internal and external information
- (v) Improves communication

Disadvantages of Intranet – Disadvantages of intranet are as follows –

- (i) Threat of sharing information and the loss of control
- (ii) Unauthorized access
- (iii) Limited bandwidth for the business
- (iv) Information overload lowers productivity
- (v) Hidden or unknown complexity and costs.

Q.4. Write down the applications and features of intranet.

Ans. Applications of Intranet – Important applications of intranet are as follows –

- (i) Sharing the detail of company rules/policies & regulations
- (ii) Access employee database
- (iii) Access product & customer data
- (iv) Sharing some common information

Intranet also use for launching personal or department-specific home pages

- (v) Submission of reports
- (vi) Corporate telephone directories.

Features of Intranet – Important features of intranet are as follows –

- (i) Intranet is fast and accurate
- (ii) Most website with large graphical images, videos, and sound, process fast on intranet
- (iii) Your firewall protects it from external threats
- (iv) It is easy to monitor with your organization
- (v) Easy communication across the company from all levels
- (vi) Helps to share ideas and discussions.

Q.5. What is Internet? Give its advantages and disadvantages.

Ans. Internet is a worldwide/global system of interconnected computer networks. It uses the standard Internet Protocol (TCP/IP). Every computer in Internet is identified by a unique IP address. IP address is a unique set of numbers (such as 110.22.33.114) which identifies a computer's location.

A special computer DNS (Domain Name Server) is used to provide a name to the IP address so that the user can locate a computer by a name. For example, a DNS server will resolve a name <https://www.shivanipublications.com> to a particular IP address to uniquely identify the computer on which this website is hosted. Internet is accessible to every user all over the world.

Advantages of Internet –

- (i) Using the Internet, you can participate in discussion groups, such as mailing lists and newsgroups.
- (ii) Allows you to send an e-mail message from every location
- (iii) The Internet is a network of computers at different locations around the world.
- (iv) Helps you to send or receive files between different computers.

Disadvantages of Internet –

- (i) The search engines may display some fake news results.
- (ii) Working on the Internet is surely tiring.
- (iii) It allows everybody to speak about everything without any limitations or censorship. That could be a bad influence on impressionable minds.
- (iv) The Internet makes us lazier – as for common things like search the nearest restaurant or finding the best hotel.
- (v) Internet could replace face to face collaborations and make us lose the human touch.

Q.6. Give the applications and features of Internet.

Ans. Applications of Internet – Important applications of Internet are as follows –

- (i) Download programs and files
- (ii) To send and receive e-mails

- (iii) Voice and video conferencing
- (iv) E-commerce
- (v) File sharing
- (vi) Browsing various types of information
- (vii) Search the web addresses for access through the search engine and chatting.

Features of the Internet – Some important features of the Internet are as follows –

- (i) A global network which connects millions of computers
- (ii) The Internet is decentralized
- (iii) Every computer over the Internet is independent
- (iv) There are various ways to access the Internet.

Q.7. Write down the differences between Internet and intranet.
Ans. Differences between Internet and intranet are as follows –

S.No.	Internet	Intranet
(i)	The Internet is a wide network of computers and is available to all.	Intranet is a network of computers designed for a certain group of users.
(ii)	Internet contains a large number of Intranets.	Intranet can be accessed from the Internet with specific restrictions.
(iii)	Number of Internet users are very high.	Number of users is limited.
(iv)	Internet contains various source of information.	Intranet only contains group-specific information.
(v)	Anyone can access the Internet	Accessible only by the organization employees or admin who have login details.
(vi)	It is not as safe as compared to Intranet.	Safe and secure network
(vii)	It is a public network	It is a private network.

Q.8. Write short note on bridges.

Ans. A bridge is a device that provides interconnection between similar networks, like LANs. Bridges operate in both the physical and the data link layers. As a physical layer device, it regenerates the signal it receives. As a data link layer device, the bridge can check the physical address (source and destination) contained in the packet. This property requires the bridge to perform minimal amount of processing. Bridges can divide a large network into smaller segments. They use the logic to keep the traffic for each segment separate. They act as a relay of frames between similar networks. They also increase the distance coverage.

A bridge operates at the data link layer giving it access to the physical addresses of all stations connected to it. When a frame enters a bridge the bridge not only regenerates the signal but checks the address of the destination and forwards the new copy only to the segment to which the address belongs. As a bridge encounters a packet, it reads the address contained in the frame and compares that address with a table of all the stations on both segments. When it finds a match, it discovers to which segment the station belongs and relays the packet only to that segment. A bridge is shown in fig. 1.4.

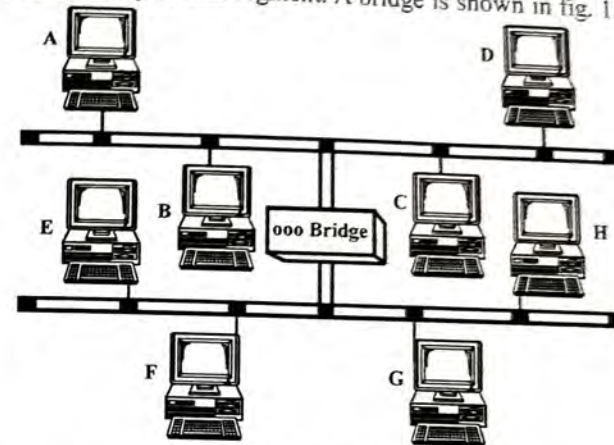


Fig. 1.4 Bridge

A bridge has filtering capability. It can check the destination address of a packet and decide if the packet should be forwarded or dropped. If the packet is to be forwarded, the decision must specify the interface. A bridge has a table that maps addresses to interfaces.

Q.9. What are the types of bridges ? Explain them.

Ans. The types of bridges are as follows –

- (i) Simple bridge
- (ii) Multiport bridge
- (iii) Transparent bridge
- (iv) Source routing bridge
- (v) Remote bridge.

(i) Simple Bridge – Simple bridges are the most primitive and least expensive type of bridge. A simple bridge links two segments and contains a table that lists the addresses of all the stations included in each of them. A simple bridge can be used an operator must sit down and enter the addresses of every station. Whenever a new station is added, the table must be modified. If a station is removed, the newly invalid address must be deleted. The logic included in a simple bridge, therefore, is of the pass/no pass variety. Installation and maintenance of simple bridges are time consuming and potentially more trouble than the cost saving are worth.

(ii) **Multiport Bridge** – A multiport bridge can be used to connect more than two LANs. Multiport bridge is shown in fig. 1.5. In this figure, the bridge has three tables, each one holding the physical addresses of stations reachable through the corresponding port.

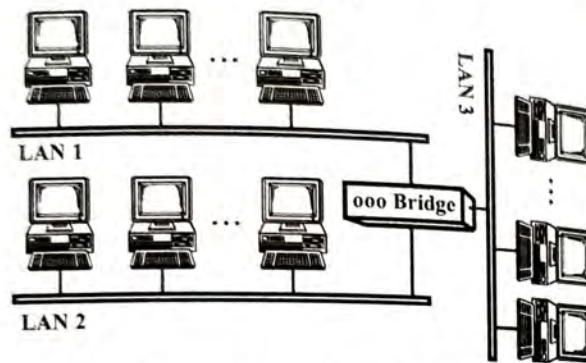


Fig. 1.5 Multiport Bridge

(iii) **Transparent Bridge** – Transparent bridge is also called *learning bridge*. A transparent bridge builds its table of station addresses on its own; it performs its bridge function. When the transparent bridge is first installed, its table is empty. As it encounters each packet, it looks at both the destination and the source addresses. It checks the destination to decide where to send the packet. If it does not yet recognize the destination address, it relays the packet to all of the stations on both segments. It uses the source address to build its table. As it reads the source address, it notes which side the packet came from and associates that address with the segment to which it belongs. Transparent bridge is shown in fig. 1.6.

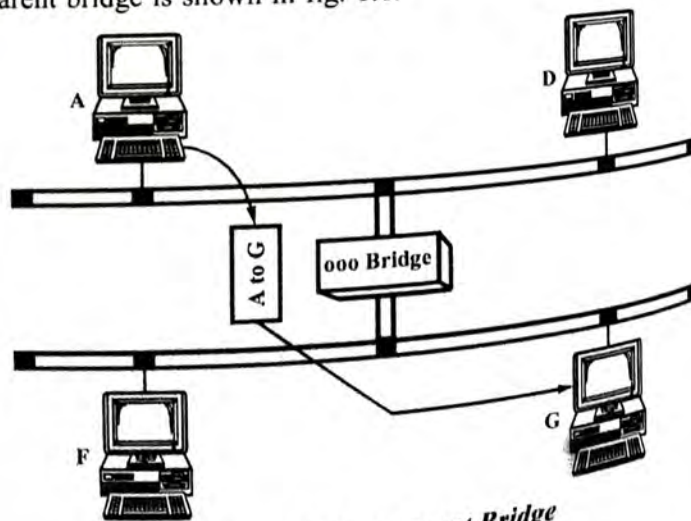


Fig. 1.6 Transparent Bridge

When station A sends its packet to station G, the bridge learns that packets coming from A are coming from the upper segment, and that station A must be located in the upper segment. Now, whenever the bridge encounters packets addressed to A, it knows to relay them only to the upper segment.

With the first packet transmitted by each station, the bridge learns the segment associated with that station. Eventually it has a complete table of station addresses and their respective segment stored in its memory.

(iv) **Source Routing Bridge** – A source routing bridge is used to network token ring LANs. In the source routing algorithm used in a bridged token ring network, the source is aware of the entire path to the destination. To the destination address, the source inserts in the packet the route that the packet should take. Thus, the intermediate nodes make no decision as to the path that the packet takes. This is the reason that the token ring bridge is called source routing bridge. The routing table can be stored on a server or in each source routing bridge. A source routing bridge network is shown in fig. 1.7.

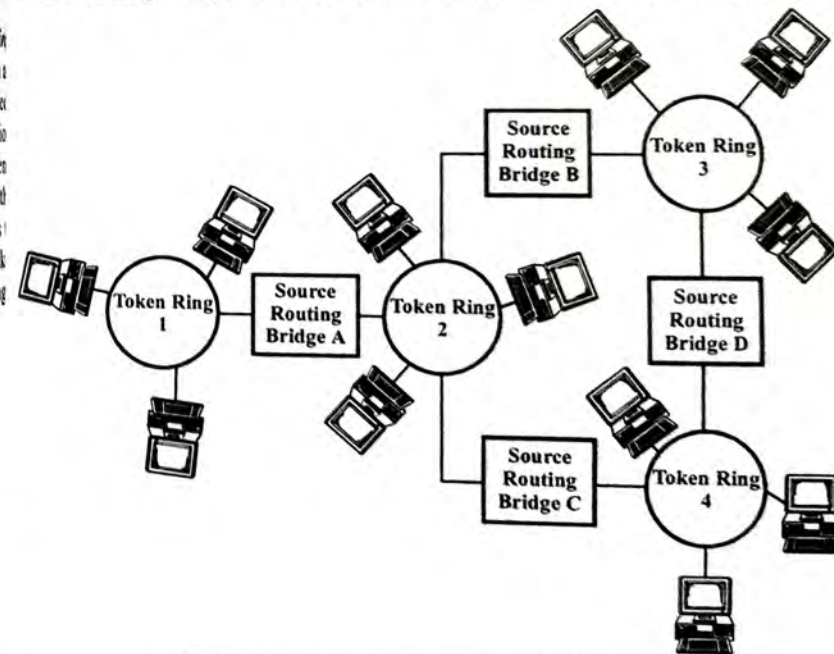


Fig. 1.7 A Source Routing Bridge Network

(v) **Remote Bridge** – The following fig. 1.8 shows a remote bridge configuration, where two bridges at remote location are linked via a WAN. WAN architecture mostly uses routers. Using remote bridges and leased dedicated telecommunication links, we can connect remote LANs.

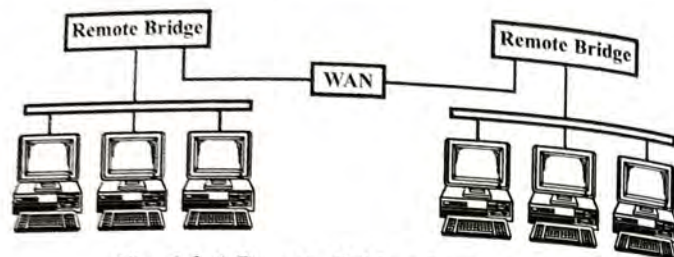


Fig. 1.8 A Remote Bridge Configuration

Bridged networks use two network routing algorithm – a spanning tree algorithm for bridging Ethernet LANs and a source routing algorithm for bridging token ring LANs.

Q.10. Write short note on routers.

Ans. A router is a device capable of interconnecting a variety of LANs and WANs, i.e., it can connect two or more networks that may or may not be similar. A router is a combination of hardware and software. The hardware consists of a network, server and a separate computer. The software consists of the operating system and the routing protocol. Routers operate in the physical, data link, and network layers.

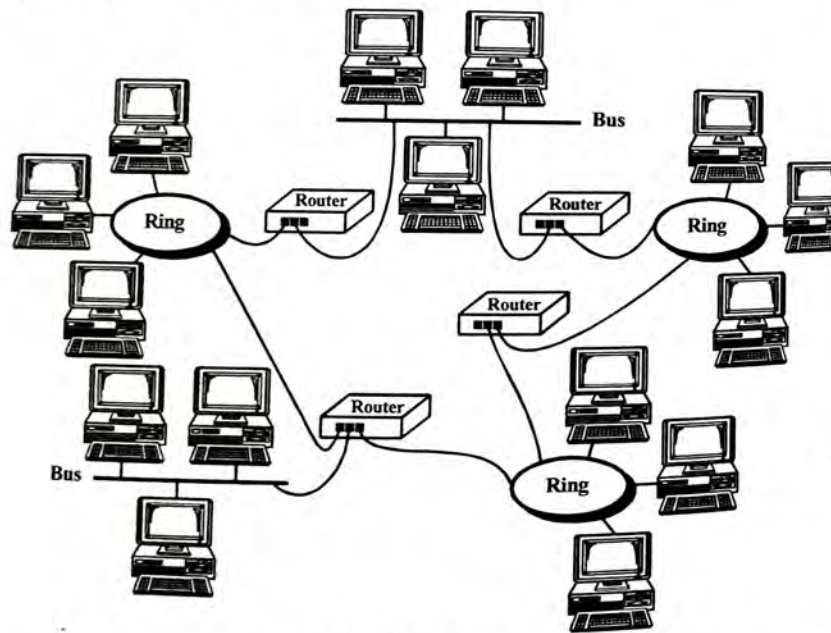


Fig. 1.9 Routers

A router, as the name indicates, routes packets through the network. Routers relay packets among multiple interconnected networks. They route

packets from one network to any of a number of potential destination networks on an Internet. Fig. 1.9 shows a possible internetwork of five networks. A packet sent from a station on one network to a station on a neighbouring network goes first to the jointly held router, which switches it over to the destination network. If there is no one router connected to both the sending and receiving networks the sending router transfers the packet across one of its connected networks to the next router in the direction of the ultimate destination. That router forwards the packet to the next router on the path, and so on, until the destination is reached.

Routers use logical and physical addressing to connect two or more logically separate networks. They do so by organizing the large network into logical network segments. Each subnetwork is given a logical address. This keeps the networks separate but they can access each other by exchanging data whenever necessary.

In short, routers make routing decisions based on the topology and conditions of the network. Routers provide reliability, good performance and great networking range, in networks.

Q.11. Give comparison between routers and bridges.

Ans. Comparison between a router and a bridge is given below –

S. No.	Router	Bridge
(i)	Router operates at network layer of OSI model.	Bridge operates at data link layer of OSI model.
(ii)	Difficult to setup and configure.	Relatively easy to configure.
(iii)	Routers are relatively expensive.	Bridges are relatively inexpensive.
(iv)	Router transmits the packets between different networks.	Bridge transmits the frame between similar networks.
(v)	Can route packets to reduce network bottlenecks.	Filter packets faster than routers.
(vi)	Routers are good solution for joining remote network.	Bridges are good for segment network.
(vii)	Join two different networks.	Extends the existing network.
(viii)	Router can accommodate multiple paths.	Bridge can accommodate single path.
(ix)	Router focuses on protocol address.	Bridge focuses on MAC address.
(x)	Routers are used in internetworking.	Bridges are used in LANs.
(xi)	Routers are both hardware and software device.	Bridges are both hardware and software device.

REVIEW OF TCP/IP PROTOCOL ARCHITECTURE – ARP/RARP, IP ADDRESSING, IP DATAGRAM FORMAT AND ITS DELIVERY, ROUTING TABLE FORMAT, ICMP MESSAGES, SUBNETTING, SUPERNETTING AND CIDR, DNS

Q.12. Write short note on TCP/IP.

(R.G.P.V., Dec. 2003)

Ans. The Transmission Control Protocol/Internet Protocol (TCP/IP) is a set of protocols that defines all transmission exchanges across the Internet. TCP is specifically designed to provide a reliable end-to-end byte stream over an unreliable internetwork. An internetwork differs from a single network because different parts may have widely different topologies, bandwidths, delays, packet sizes and other parameters. TCP is designed to dynamically adapt properties of the internetwork and to be robust in the face of many kinds of failures.

TCP is a connection-oriented protocol and IP is a connectionless datagram protocol in the transport layer. These services provided by TCP are richer than those provided by IP. The TCP provides full transport layer services to applications. A virtual circuit is established between sender and receiver before transmission. A TCP entity at one machine accepts user data streams from local processes, breaks them up into pieces and sends each piece as a separate IP datagram. When IP datagrams containing TCP data arrive at the destination machine, the TCP entity reconstructs the original byte streams.

The IP layer gives no guarantee to deliver the datagram properly so it is up to the TCP to timeout and retransmit if necessary. Also TCP reassembles the byte streams into proper sequence.

In short, TCP gives the reliability which the IP does not provide. Reliability means to handle flow control, error control and loss control. TCP was initially defined in RFC 793.

Internet Protocol (IP) is the transmission mechanism used by the TCP/IP protocol. IP itself has four supporting protocols – ARP, RARP, ICMP and IGMP. IP is an unreliable and connectionless datagram protocol, a best effort delivery service. It provides no error control. When reliability is important, IP is coupled with a reliable protocol, TCP. IP transports data in packets called datagrams. IP is used when speed is more important than reliability.

Q.13. Describe the function of each layer of TCP/IP reference model.

(R.G.P.V., Dec. 2009)

Ans. The ARPANET was a research network sponsored by the DoD (U.S. Department of Defense). It eventually connected hundreds of universities and government installations using leased telephone lines. When satellite and

radio networks were added later, the existing protocols had trouble interworking with them, so a new reference architecture was needed. Thus, the ability to connect multiple networks together in a seamless way was one of the major design goals from the very beginning. This architecture later became known as the **TCP/IP reference model** after its two primary protocols.

The TCP/IP protocol suite was developed prior to the OSI model. Therefore, the layers in the TCP/IP protocol suite do not exactly match those in the OSI model. The original TCP/IP protocol suite was defined as having four layers – host-to-network, internet, transport and application.

Besides the OSI model another model which is in practical use is the TCP/IP model which is five layer model as shown in fig. 1.10. Internet services generally support TCP/IP protocol. The functions of various layers are discussed below –

Application Layer
Transport Layer
Internet Layer
Network Access Layer
Physical Layer

Fig. 1.10 The TCP/IP Reference Model

(i) **Physical Layer** – It covers the physical interface between a data transmission device (e.g., workstation, computer) and a transmission medium or network. This layer is concerned with specifying the characteristics of the transmission medium, the nature of the signals, the data rate, and related matters.

(ii) **Network Access Layer** – The network access layer is concerned with the exchange of data between an end system and the network to which it is attached. The sending computer must provide the network with the address of the destination computer, so that the network may route the data to the appropriate destination. The sending computer may wish to invoke certain services, such as priority that might be provided by the network. The specific software used at this layer depends on the type of network to be used-different standards have been developed for circuit-switching, packet-switching, local area networks and others. This protocol is not defined and varies from host-to-host and network-to-network.

(iii) **Internet Layer** – All these requirements led to the choice of a packet-switching network based on a connectionless internetwork layer. This layer, known as internet layer, is the linchpin that holds the whole architecture together. The internet layer defines an official packet format and protocol called **IP (Internet Protocol)**. The internet protocol is used at this layer to provide the routing function across multiple networks. This protocol is implemented not only in the end systems but also in the routers. A router is a processor that connects two networks and whose primary function is to relay data from one network to the other on its route from the source to the destination end system.

(iv) **Transport Layer** – This layer is designed to permit peer entities on the source and destination hosts to carry on a conversation, just as in the OSI transport layer. Two end-to-end transport protocols have been defined here.

The first one, TCP (Transmission Control Protocol) is a reliable connection-oriented protocol that allows a byte stream originating on one machine to be delivered without error on any other machine in the Internet. It fragments the incoming byte stream into discrete messages and passes each one on to the internet layer. TCP also handles flow control to make sure a fast transmitter cannot swamp a slow receiver with more messages than it can handle.

The second protocol in this layer, UDP (User Datagram Protocol) is an unreliable connectionless protocol for application that do not want TCP's sequencing or flow control and wish to provide their own. It is also widely used for one-shot, client-server type request-reply queries and applications in which, prompt delivery is more important than accurate delivery such as transmitting speech or video.

(v) **Application Layer** – On top of the transport layer is the application layer. It contains all the higher-level protocols. The early ones included virtual terminal (TELNET), file transfer (FTP), and electronic mail (SMTP). The virtual terminal protocol allows a user on one machine to log onto a distant machine and work there. The file transfer protocol provides a way to move data efficiently from one machine to another. Electronic mail was originally just a kind of file transfer, but later a specialized protocol (SMTP) was developed for it. Many other protocols have been added to these over the years – the Domain Name System (DNS) for mapping host names onto their network addresses, NNTP, the protocol for moving USENET news articles around, and HTTP.

Q.14. What is ARP ?

Ans. The ARP (address resolution protocol) is a protocol in the network layer. It is supported by TCP/IP. The ARP associates an IP address with its physical address. On a typical physical network, such as a LAN, each device on a link is identified by a physical or station address usually imprinted on the network interface card (NIC).

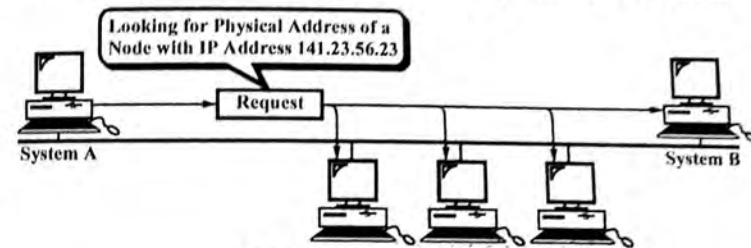
Physical address have local jurisdiction and can be changed easily. For example, if the NIC on a particular machine fails, the physical address changes. On the otherhand, the IP addresses have universal jurisdiction and cannot be changed. ARP is used to find the physical address of the node when its internet address is known.

Q.15. How does the address resolution protocol (ARP) work ?

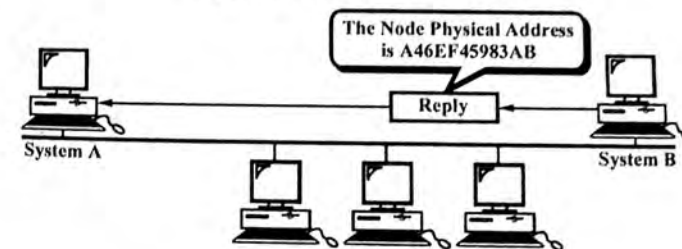
Ans. Anytime a host, or a router, needs to find the physical address of another host on its network, it sends an ARP query packet that includes the physical and IP addresses of the sender and the IP address of the receiver. Since the sender does not know the physical address of the receiver, the query is broadcast over the network (see fig. 1.11). Every host or router on the

network receives and processes the ARP query packet, but only the intended recipient recognizes its IP address and sends back an ARP response packet. The response packet contains the recipient's IP and physical addresses. The packet is unicast directly to the inquirer using the physical address received in the query packet.

In fig. 1.11 (a), the system on the left (A) has a packet that needs to be delivered to another system (B) with IP address 141.23.56.23. System A needs to pass the packet to its data link layer for the actual delivery, but it does know the physical address of the recipient (i.e. system B). It uses the services of ARP by asking the ARP protocol to send a broadcast ARP request packet to ask for the physical address of a system with an IP address of 141.23.56.23.



(a) ARP Request is Broadcast



(b) ARP Reply is Unicast

Fig. 1.11 ARP Operation

This packet is received by every system on the physical network, but only system B will answer it, as shown in fig. 1.11 (b). System B sends an ARP reply packet that includes its physical address. Now, the system A can send all the packet it has for this destination using the physical address obtained.

The ARP provides a mechanism that is both reasonably efficient and easy to maintain.

Q.16. List the various steps involved in an ARP process.

Ans. The steps involved in an ARP process are as follows –

- (i) The sender knows the IP address of the target.

(ii) IP asks ARP to create an ARP request message, filling in the sender physical address, the sender IP address, and the target IP address. The target physical address field is filled with 0s.

(iii) The message is passed to the data link layer where it is encapsulated in a frame using the physical address of the sender as the source address and the physical broadcast address as the destination address.

(iv) Each host or router receives the frame. Since the frame contains a broadcast destination address, all stations remove the message and pass it to ARP. All machines except the one targeted drop the packet. The target machine recognizes the IP address.

(v) The target machine replies with an ARP reply message that contains its physical address. The message is unicast.

(vi) The sender receives the reply message. Thus, it knows the physical address of the target machine.

(vii) Now, the IP datagram, which carries data for the target machine, is encapsulated in a frame and is unicast to the destination.

Q.17. Show how the address resolution protocol is used to dynamically map a logical address to a physical address.

Ans. A solution to the address resolution problem for networks such as the Ethernet has been found by the designers of TCP/IP protocols. This solution permits new hosts or routers to be added to the network without recompiling code. Also, it does not need maintenance of a centralized database. A low-level protocol is used to bind addresses dynamically that avoids maintaining a table of mappings. The ARP gives a mechanism that is both easy and efficient.

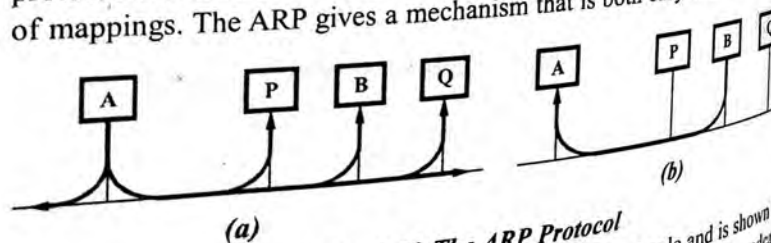


Fig. 1.12 The ARP Protocol

The idea behind dynamic resolution with ARP is simple and is shown in fig. 1.12. If host A wants to resolve IP address I_B , then a packet is broadcasted by host A to ask to host with IP address I_B to respond with its physical address, P_B . The request is received by all hosts, but only host B identifies its IP address and send a response that has its physical address. This response is received by host A and uses the physical address to send the internet packet directly to B.

Q.18. Briefly describe ARP message format.

Ans. Fig. 1.13 shows the format of an ARP packet.

Hardware Type		Protocol Type
Hardware Length	Protocol Length	Operation Request 1, Reply 2
Sender Hardware Address (For example, 6 bytes for Ethernet)		
Sender Protocol Address (For example, 4 bytes for IP)		
Target Hardware Address (For example, 6 bytes for Ethernet)		
Target Protocol Address (For example, 4 bytes for IP)		

Fig. 1.13 ARP Packet

The fields are discussed as follows –

(i) **HTYPE (Hardware Type)** – It is a 16-bit field defining the type of the network on which ARP is running. Each LAN has been assigned an integer based on its type. For example, Ethernet is given the type 1. ARP can be used on any physical network.

(ii) **PTYPE (Protocol Type)** – It is a 16-bit field defining the protocol. For example, the value of this field for the IPv4 protocol is 0800_{16} . ARP can be used with any higher-level protocol.

(iii) **HLEN (Hardware Length)** – It is an 8-bit field defining the length of the physical address in bytes. For example, for Ethernet the value is 6.

(iv) **PLEN (Protocol Length)** – It is an 8-bit field defining the length of the logical address in bytes. For example, for the IPv4 protocol the value is 4.

(v) **OPER (Operation)** – It is a 16-bit field defining the type of packet. Two types of packet are defined – ARP request (1), ARP reply (2).

(vi) **SHA (Sender Hardware Address)** – It is a variable-length field defining the physical address of the sender. For example, for Ethernet this field is 6 bytes long.

(vii) **SPA (Sender Protocol Address)** – It is a variable-length field defining the logical address of the sender. For the IP protocol, this field is 4 bytes long.

(viii) **THA (Target Hardware Address)** – It is a variable-length field defining the physical address of the target. For example, for Ethernet this field is 6 bytes long. For an ARP request message, this field is all 0s because the sender does not know the physical address of the target.

(ix) **TPA (Target Protocol Address)** – It is a variable-length field defining the logical address of the target. For the IPv4 protocol, this field is 4 bytes long.

Q.19. What is RARP (Reverse Address Resolution Protocol)? How does it work?

Ans. RARP protocol is the part of network layer protocols, which is supported by TCP/IP. RARP finds the logical (Internet) address for a machine that only knows its physical address.

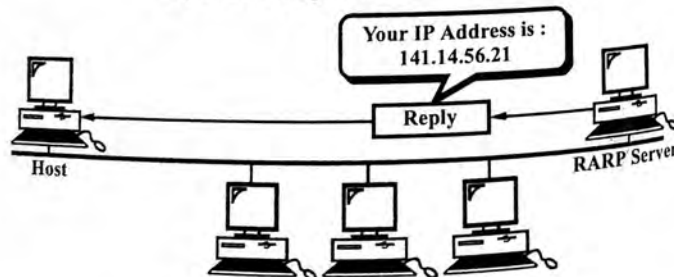
Each host or router is assigned one or more logical (IP) addresses, which are unique and independent of the physical address of the machine. A host or a router must need to know its own IP address or addresses to create an IP datagram. The IP address of a machine is usually read from its configuration file stored on a disk file.

However, a diskless machine is usually booted from ROM. The ROM is installed by the manufacturer. It cannot include the IP address, since the IP addresses are assigned by the network administrator on a network.

The machine can get its physical address by reading its NIC, which is unique locally. It can then use the physical address to get the logical address using the RARP protocol. A RARP request is created and broadcast on the local network. Another machine on the local network that knows all the IP addresses will respond with a RARP reply. The requesting machine must be running a RARP client program, and the responding machine must be running a RARP server program as shown in fig. 1.14.



(a) RARP Request is Broadcast



(b) RARP Reply is Unicast

Fig. 1.14 RARP Operation

In fig. 1.14 (a), the diskless host on the left is booted. To get its IP address, it broadcasts a RARP request to all systems on the network. This packet is received by every host or router on the physical network, but only the RARP server on the right will answer it, as shown in fig. 1.14 (b). The server sends a RARP reply packet that includes the IP address of the requestor.

Q.20. Explain the RARP packet format.

Ans. The format of RARP packet is exactly same as ARP packet except the operation field is either 3 RARP request or 4 RARP reply. Fig. 1.15 shows the RARP packet format.

Hardware Type		Protocol Type
Hardware Length	Protocol Length	Operation Request 3, Reply 4
Sender Hardware Address (For Example, 6 bytes for Ethernet)		
Sender Protocol Address (For Example, 4 bytes for IP) (It is not Filled for Request)		
Target Hardware Address (For Example, 6 bytes for Ethernet) (It is not Filled for Request)		
Target Protocol Address (For Example, 4 bytes for IP) (It is not Filled for Request)		

Fig. 1.15 RARP Packet Format

Refer to Q.18, for the description of each field.

Q.21. Discuss the concept of IP addressing.

Ans. Internet is a virtual structure, imagined by its designers and implemented entirely in software. Thus, the designers are free to choose packet formats and sizes, address delivery techniques and so on, nothing is dictated by hardware. For addresses, the designers of TCP/IP choose a scheme analogous to physical network addressing in which each host on the Internet is assigned a 32-bit integer address called its IP address.

An IP address is a unique number which helps to identify each and every computer, node or any resource like the dial-in-modem on the Internet. It is called personally identifiable information that is automatically captured by another computer when any communication link is established over the Internet. IP address is important in terms of security through which unauthorized user can be traced. The IP address is divided into 2 parts. The prefix of an IP address identifies a network and suffix part identifies the particular host on the given network. This has created a two level addressing hierarchy as shown in fig. 1.16.

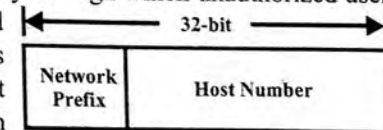


Fig. 1.16 Two Level Address Hierarchy

The IP addresses are expressed in the form of Binary, Dotted decimal and Hexadecimal. In the dotted decimal notation the IP address is divided into four parts each separated by a dot, 140.30.46.2.

Each of the four sections in the IP address reveals a part of the address. Each part is called octet as each octet can be represented by binary 8 bits and can represent a decimal number from 0-255.

Using combinations of these octets, millions of addresses can be generated. IP addresses are organized from left to right, where the extreme left hand octet describes the largest network while the right hand octet describes the actual network.

Q.22. Give the IP address or range of IP address which is reserved for –

- | | |
|-------------------------|--------------------------|
| (i) Broadcasting | (ii) Multicasting |
| (iii) Unicasting | (iv) Future use. |

Ans. (i) Broadcasting – IP reserves the entire range of addresses from 255.0.0.0 through 255.255.255.255 for broadcast and this range should not be considered part of the normal class E range.

(ii) Multicasting – Class D addresses are used for multicasting. The multicast addresses are in the range 224.0.0.0 through 239.255.255.255.

(iii) Unicasting – The classes A, B, and C addresses were designed for unicasting. The class A addresses are in range 0.0.0.0 to 127.255.255.255. The class B addresses are in range 128.0.0.0 to 191.255.255.255. The class C addresses are in range 192.0.0.0 to 223.255.255.255.

(iv) Future Use – The class E addresses were reserved for future use. It falls in the range 240.0.0.0 through 255.255.255.255.

Q.23. Give the advantages and disadvantages of IP address.

Ans. Advantages of IP Address –

- IP addresses are unique numbers, i.e., no two systems directly connected to the Internet can have the same IP address.
- IP addresses are standardized and follow global conventions. All the networks that are given by Internet NIC under the authority of Assigned Numbers Authority (IANA) to the ISPs who allocate the host portion of the IP address to the machines that are connected to the network only. Only real Internet hosts IP addresses.

Disadvantages of IP Address –

- They are very difficult to remember and the possibility of committing an error is quite high.
- We are running short of addresses as the network grow constantly.
- If a computer is transferred from one place to another, it is quite possible its IP addresses will get modified thus causing inconvenience.
- Routing problem has arisen because the number of resources on the Internet have increased and so has the data flow.

Q.24. How can we find the subnet address from given IP address ?

Ans. Given an IP address, we can find the subnet address the same way we found the network address. We apply mask to the address. We can do this in two ways – straight or short-cut.

(i) Straight Method – In the straight method, we use binary notation for both the address and the mask and then apply the AND operation to find the subnet address.

Suppose, the destination address is 200.45.34.56 and the subnet mask is 255.255.240.0. Now we want to find the subnetwork address. Then, we apply the AND operation on the address and the subnet mask.

Address	→ 11001000	00101101	00100010	00111000
Subnet Mask	→ 11111111	11111111	11110000	00000000
Subnetwork Address	→ 11001000	00101101	00100000	00000000

The subnetwork address is 200.45.32.0

(ii) Short-cut Method – If the subnet mask is contiguous, we can use a short-cut as we did previously to find the network address using the default mask. The three rules we use are –

- If the byte in the mask is 255, copy the byte in the address.
- If the byte in the mask is 0, replace the byte in the address with 0.
- If the byte in the mask is neither 255 nor 0, we right the mask and the address in binary and apply the AND operation.

Q.25. Discuss IP datagram format in detail. (R.G.P.V., Dec. 2004, 2005)

Ans. On a physical network the unit of transfer is a frame that contains a header and data, where the header gives information such as the source and destination addresses. The internet calls its basic transfer unit an *Internet datagram*, sometimes referred to as an *IP datagram* or a *datagram*. That is packets in the IP layer are called datagrams.

A datagram is divided into header and data areas. Also, they contains the source and destination addresses and a type field that identifies the contents of the datagram. A general form of a datagram is shown in fig. 1.17.

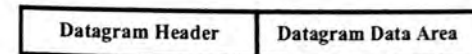


Fig. 1.17 General Form of an IP Datagram

Fig. 1.18 shows the arrangement of fields in a datagram. The normal size of IP header is 20 to 60 bytes. The brief description of each field is given below –

(i) Version (VER) – This 4-bit field defines the version of the IP protocol. Currently the version is 4, However, version 6 may replace version 4 in a few years. This field tells the IP software running in the processing machine that the datagram has the format of version 4.

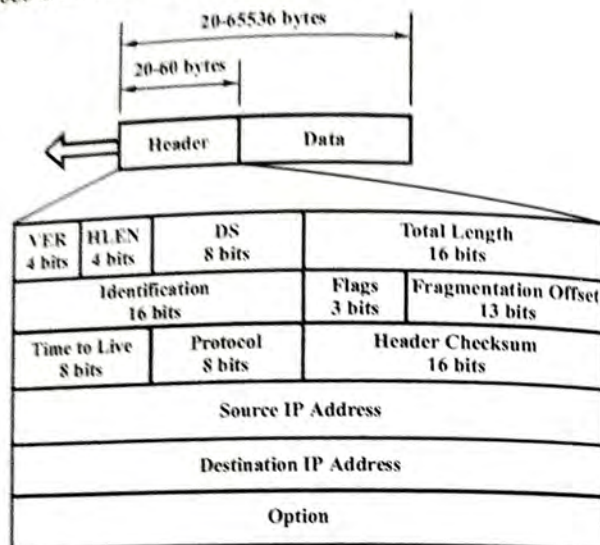


Fig. 1.18 IP Datagram Format

(ii) **Header Length** – This 4-bit field defines the length of the header in multipliers of four bytes. The four bits can represent a number between 0 and 15, which when multiplied by 4, gives a maximum of 64 bytes.

(iii) **Differentiated Services (DS)** – This 8-bit field, previously called service type, is now called differentiated services Fig. 1.19 shows both interpretations.

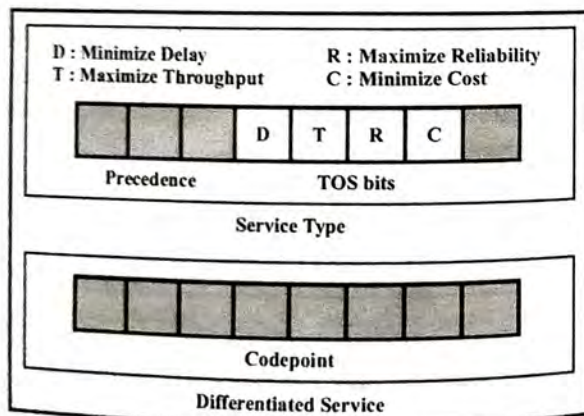


Fig. 1.19 Service Type of Differentiated Services

(a) **Service Type** – In this field, the first 3 bits are called precedence bits and next 4 bits are called TOS bits and the last bit is not used.

(1) **Precedence** – It is an 3-bit subfield ranging from 0 (000 in binary) to 7 (111 in binary). It defines the priority of the datagram during

congestion. During congestion, those datagrams having lowest priority are discarded first.

(2) **TOS bits** – It is a four bit subfield with each bit having special meaning. The bits and their interpretation are shown in table 1.1.

Table 1.1 Types of Service

TOS Bits	Description
0000	Normal (default)
0001	Minimum cost
0010	Maximum reliability
0100	Maximum throughput
1000	Minimize delay

(b) **Differentiated Services** – In this field, the first 6 bits make up the code points subfield and the last two bits are not used. The codepoint field is used in two different ways –

(1) When three rightmost bits are 0's, the 3 leftmost bits are interpreted the same as the precedence bits in the service type interpretation.

(2) When 3 rightmost bits are all not 0's, the 6 bits define 64 services on the priority assignment by the Internet according to table 1.2.

Table 1.2 Value of Codepoints

Category	Codepoint	Assigning Authority
1	XXXXX0	Internet
2	XXXX11	Local
3	XXXX01	Temporary or experimental

(iv) **Total Length** – This 16-bit field defines the total length (header plus data) of the IP datagram in bytes. This is a two-byte field and can define upto 65,535 bytes.

Header length can be found by multiplying the value on the HLEN field by four.

Length of data = Total length – Header length

(v) **Identification** – This field is used in fragmentation.

(vi) **Fragmentation Offset** – This is a 13 bit pointer that shows the offset of the data in the original datagram.

(vii) **Flags** – The 3 bits in the flags field deal with fragmentation.

(viii) **Time-to-live** – This field defines the number of hops a datagram can travel before it is discarded. The source host, when it creates the datagram, sets this field to an initial value. Then as the datagram travels through the Internet, router by router, each router decrements this value by 1. If this value becomes 0 before the datagram reaches its final destination, the datagram is discarded. This prevents a datagram from back and forth forever between routers.

Source IP Address – The source address field is a four-byte field that identifies the original source of the datagram.

(x) **Source IP Address** – The source address field is a four-byte (32-bit) Internet address. It identifies the original source of the datagram.

(xi) **Header Checksum** – This field is used to check the integrity of the header, not the rest of the packet.

(xii) **Destination IP Address** – This is a four byte field used to identify the final destination of the datagram.

(xiii) **Options** – This field gives more functionality to the IP datagram. This can carry field that control routing, timing, management and alignment.

Q.26. What is the purpose of time-to-live field in IP datagram header?
Or

Q.26. What is the purpose of the time-to-live field in an IP header?
(R.G.P.V., June 2004, Dec. 2004, 2007)

Ans. The time-to-live field in IP header specifies how long, in seconds, the datagram is allowed to remain in the Internet system. The idea is simple and important when a computer injects a datagram to the Internet, it sets a maximum time that the datagram should survive. Routers and hosts that process the datagram must decrement the time-to-live field as time passes and remove the datagram from the Internet when its time expires.

Calculating the exact time is difficult because usually routers do not know the transit time for the physical networks. There are some rules which are followed to simplify the processing and make it easy to handle datagram without synchronizing the clocks.

First, each router along the path from source to destination is required to decrement the time-to-live field by 1 when it processes the datagram header. To handle the case of overloaded routers which introduce long delays, each router records the local time when the datagram arrives and decrements the time-to-live by the number of seconds the datagram remained inside the router waiting for service.

Whenever this field reaches to zero, router discards the datagram and sends an error message back to the source. This process guarantees that the datagram cannot travel around an internet forever, even if routing tables become corrupt and routers route datagrams in a circle.

Q.27. What is direct and indirect delivery of IP packets ?
Ans. If...

Ans. If final destination (host) of the packet is connected to the same physical network as the deliverer then the delivery of the packet within same network is known as **direct delivery** (see fig. 1.20).

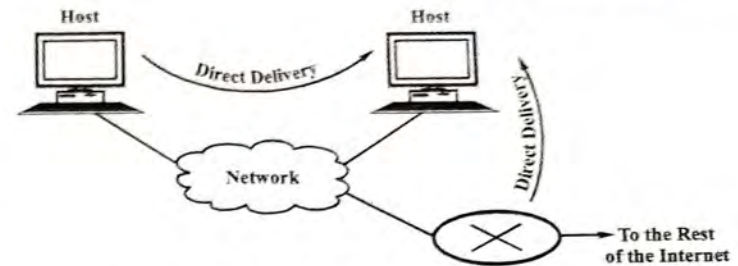


Fig. 1.20 Direct Delivery

The sender can easily determine whether it is direct or indirect delivery by extracting the network address of the destination packet and match this address with the addresses of the networks to which it is connected. If it matches the delivery is direct otherwise it is indirect.

If the destination host is not on the same network as the deliverer, the packet is delivered indirectly and delivery is known as **indirect delivery**. In indirect delivery packet goes from router to router until it reaches to final destination.

It is noted that a delivery is always involves one direct delivery and other may or may not indirect deliveries. Also, the last delivery is always a direct delivery. See in fig. 1.21.

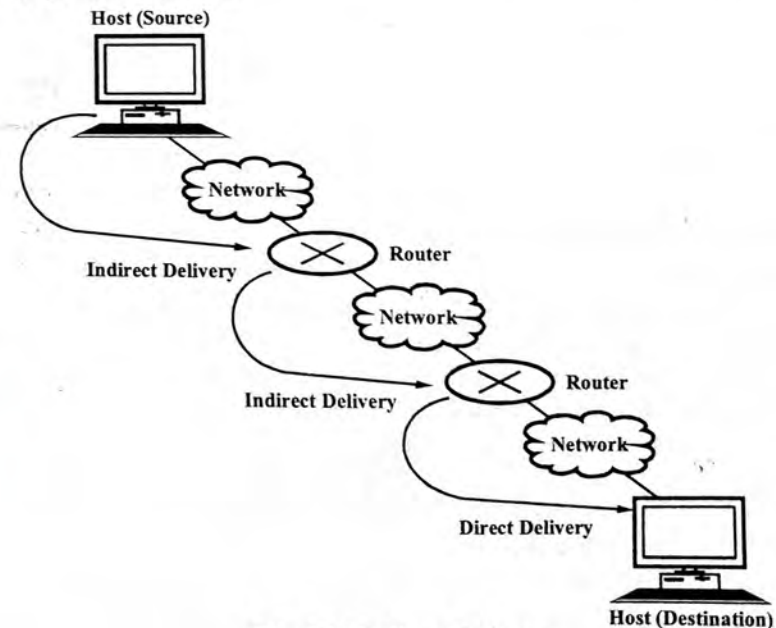


Fig. 1.21 Indirect Delivery

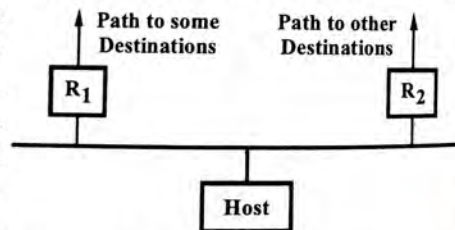
Q.28. What is IP routing ? Explain.

Ans. In a packet switching system, *routing* refers to the process of choosing a path over which to send packets, and *router* refers to a computer making the choice. Routing occurs at several levels. For example, within a WAN, the network itself is responsible for routing packets from the time they enter until they leave. This internal routing is completely self-contained inside the wide area network. Machines on the outside cannot participate in decisions they merely view the network as an entity that delivers packets.

The process of finding the route (next hop) for a datagram is known as IP forwarding or Internet routing or **IP routing**. The information used to make routing decisions is known as *IP routing information*. Like routing within a single network, IP routing chooses a path over which a datagram should be sent. Unlike routing within a single network, the IP routing algorithm must choose how to send a datagram across multiple physical networks.

Routing in an Internet can be difficult, especially among computers that have multiple physical network connections. Ideally the routing software would examine network load, datagram length or the type of service specified in the datagram header when selecting the best path. Most Internet routing software is much less sophisticated, however, and selects routes based on fixed assumptions about shortest paths.

An Internet is composed of multiple physical networks interconnected by computers called routers. Each router has direct connections to two or more networks. By contrast, a host computer usually connects directly to one physical network. It is also possible, we may have a multihomed host connected directly to multiple networks.

**Fig. 1.22 IP Routing**

Both host and routers participate in routing an IP datagram to its destination. When an application program on a host attempts to communicate, the TCP/IP protocols eventually generate one or more IP datagrams. The host must make an initial routing decision when it chooses where to send the datagrams. Fig. 1.22 shows that hosts must make routing decisions even if they have only one network connections. The primary purpose of routers is to make IP routing decisions. However computer with multiple network connections can act as router.

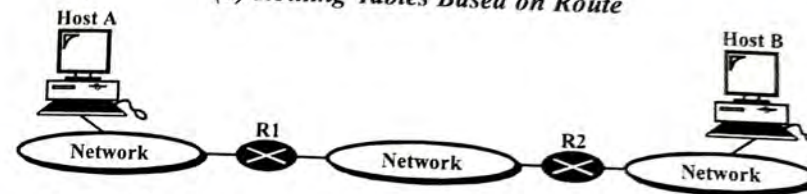
Q.29. Explain the various IP routing techniques for the Internet.

Ans. Routing means finding the route for a datagram. Routing requires a host or a router to have a routing table. When a host has a packet to send or

when a router has received a packet to be forwarded, it looks at this table to find the route to the final destination. Some of routing techniques are as follows –

(i) **Next-hop Routing** – A technique which reduces the contents of a routing table is called next-hop routing. In this technique, the routing table holds only the address of the next hop instead of holding information about the complete route. The entries of a routing table must be consistent with each other. Fig. 1.23 shows how routing tables can be simplified using this technique.

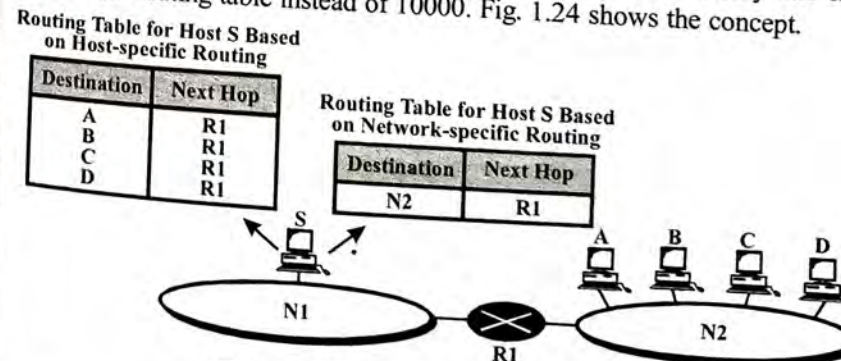
Routing Table for Host A		Routing Table for R1		Routing Table for R2	
Destination	Route	Destination	Route	Destination	Route
Host B	R1, R2, Host B	Host B	R2, Host B	Host B	Host B

(a) Routing Tables Based on Route

Routing Table for Host A		Routing Table for R1		Routing Table for R2	
Destination	Next Hop	Destination	Next Hop	Destination	Next Hop
Host B	R1	Host B	R2	Host B	—

(b) Routing Tables Based on Next Hop**Fig. 1.23 Next-hop Routing**

(ii) **Network-specific Routing** – A second technique to reduce the routing table and simplify the searching process is known as network-specific routing. Here, instead of having an entry for every host connected to the same physical network, we have only one entry to define the address of the network itself. In other words, we treat all hosts connected to the same network as one single entity. For instance, if 10000 hosts are attached to the same network, only one entry exists in the routing table instead of 10000. Fig. 1.24 shows the concept.

**Fig. 1.24 Network-specific Routing**

(iii) **Host-specific Routing** – In this technique the destination host address is given in the routing table. The idea of host-specific routing is just the opposite to the network-specific routing. Here efficiency is sacrificed for other advantages. Although it is not efficient to put the host address in the routing table, there are occasions in which the administrator wants to have more control over routing. For example, in fig. 1.25 if the administrator wants all packets arriving for host B delivered to router R3 instead of R1, one single entry in the routing table of host A can explicitly define the route.

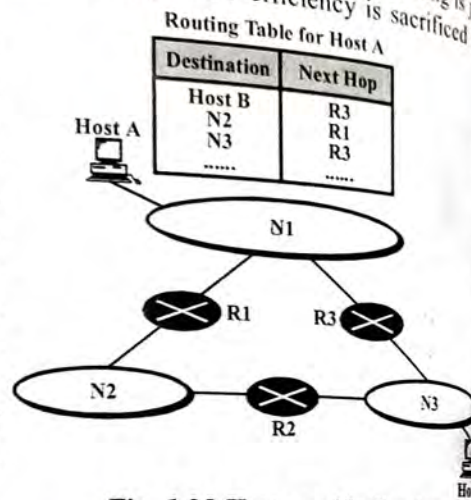


Fig. 1.25 Host-specific Routing

(iv) **Default Routing** – Another technique to simplify routing is default routing. In fig. 1.26 host A is connected to a network with two routers. Router R1 is used to route the packets to hosts connected to network N2. However, for the rest of the Internet, router R2 is used. Therefore, instead of listing all the networks in the entire Internet, host A can just have one entry called default (network address 0.0.0.0).

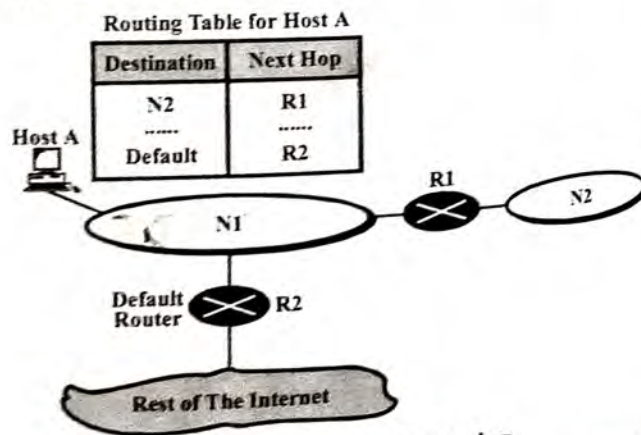


Fig. 1.26 Default Routing

(v) **Table Driven IP Routing** – The usual IP routing algorithm employs an Internet routing table on each machine that stores information about possible destinations and how to search them. Because both hosts and routers

both have IP routing tables. Whenever the IP routing software in a host or router needs to transmit a datagram, it consults the routing table to decide where to send the datagram.

If every routing table contained information about every possible destination address, it would be impossible to keep the tables current. Furthermore, because the number of possible destinations is large, machines would have insufficient space to store the information.

Conceptually, we would like to use the principle of information hiding and allow machines to make routing decisions with minimal information. For example, we would like to isolate information about specific hosts to the local environment in which they exist and arrange for machines that are far away to route packets to them without knowing such details. Fortunately, the IP address scheme helps to achieve this goal. IP addresses are assigned to make all machines connected to a given physical network share a common prefix. Such an assignment makes the test for direct delivery efficient. It also means that routing tables only need to contain network prefixes and not full IP addresses.

Q.30. Write short note on ICMP. (R.G.P.V., Dec. 2004, June 2008)

Ans. The Internet Control Message Protocol allows routers to send error or control messages to other routers or hosts. ICMP provides communication between the Internet protocol software on one machine and the Internet protocol software on another machine. Technically, ICMP is an error reporting mechanism. It provides a way for routers that encounter an error to report the error to the original source.

The ICMP has been designed to compensate for the two deficiencies, the lack of error control and lack of assistance mechanisms for host and management queries, in IP protocol. It is a companion to the IP protocol. ICMP itself is a network layer protocol. However, its messages are not passed directly to the data link layer as would be expected. Instead, the messages are first encapsulated inside IP datagram before going to the lower layer (see fig. 1.27). The value of the protocol field in the IP datagram is 1 to indicate that the IP data is an ICMP message.

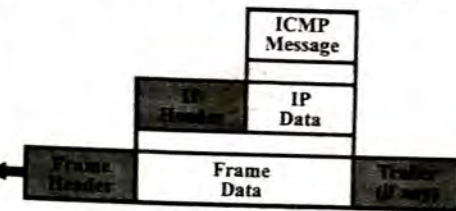


Fig. 1.27 ICMP Encapsulation

Q.31. Discuss the ICMP message format.

Ans. All ICMP messages begin with an eight-byte header followed by a variable-length data field. Although the general format of the header is different for each message type, the first 4 bytes are common to all. As fig. 1.28

shows, the first field, ICMP type, defines the type of the message. The code field specifies the reason for the particular message type. The last common field is the checksum field. The rest of the header is specific for each message type.

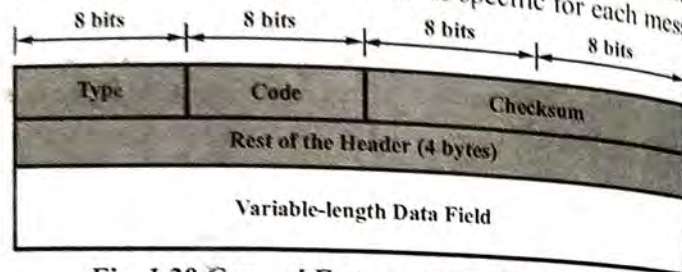


Fig. 1.28 General Format of ICMP Messages

The data field in error messages carries information for finding the original packet that had the error. In query messages, the data field carries extra information based on the type of the query.

Q.32. Describe the various ICMP message types.

Ans. ICMP messages are divided into broad categories – error-reporting messages and query messages as shown in fig. 1.29.

The error-reporting messages report problems that a router or a host (destination) may encounter when it processes IP packet.

The query messages, which occur in pairs, help a host or a network manager get specific information from a router or another host. For example, nodes can discover their neighbours. Also, hosts can discover and learn about routers on their network and routers can help a node redirect its messages.

Table 1.3 lists the ICMP messages in each category.

Table 1.3 ICMP Messages

Category	Type	Message
Error-reporting messages	3	Destination unreachable
	4	Source quench
	11	Time exceeded
	12	Parameter problem
Query messages	5	Redirection
	8 or 0	Echo request or reply
	13 or 14	Timestamp request or reply
	17 or 18	Address mask request or reply
	10 or 9	Router solicitation or advertisement

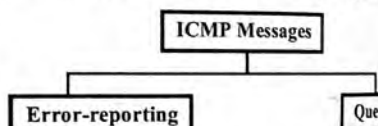


Fig. 1.29 ICMP Messages

Q.33. Explain some typical control messages sent by ICMP.

Ans. About a dozen types of ICMP messages are defined. The following list explains some typical control messages sent by ICMP.

(i) **Destination Unreachable** – As we know that IP cannot guarantee delivery of a packet. The destination may not exist or it may be down, the sender may have made a source route request that cannot be carried out or a packet with its do not fragment bit set may be too large to be encapsulated into a frame. In such cases the router detecting the error sends an ICMP packet to the original sender.

(ii) **Echo Request** – ICMP uses this packet to determine whether a particular destination is reachable. For example, if A wants to know whether B is reachable it sends an echo request packet addressed to B. If B receives the packet it responds by sending an echo reply packet back to A. The echo reply packet will return any data placed in the echo request packet.

(iii) **Echo Reply** – Sent in response to an echo request.

(iv) **Redirect** – Suppose a host station sends a packet to a router and the router knows the packet could have been delivered faster via some other router. Router sends a redirect packet back to the host for future routing. It informs the host where the other router is and that it should send future packets with the same destination to it. This allows the host to update its routing tables dynamically and to take advantage of changing conditions in the network. The redirect packet is not used for router-to-router router updates because IP packets contain the source address of the most recent router that had it.

(v) **Parameter Problem** – Suppose an IP packet contains an error or an illegal value in one of its header fields. A router discovering the error sends a parameter problem packet back to the source. This packet contains the IP header in question and a pointer to the header field that is in error.

(vi) **Time Exceeded** – A time exceeded packet is sent when the time-to-live field in an IP packet reaches 0 or when the reassemble timer expires.

(vii) **Source Quench** – This message was formerly used to throttle hosts that were sending too many packets. When a host received this message, it was expected to slow down. It is rarely used because when congestion occurs, these packets tend to add more fuel to the fire. Congestion control in the Internet is now done largely in the transport layer.

(viii) **Timestamp Request and Reply** – Timestamp packets allow a host to estimate the time required for a round trip between it and another host. A host creates and sends a timestamp request packet containing the time of transmission (original timestamp). When the receiving host gets the packet it creates a timestamp reply packet that contains the original timestamp, the time

at which the receiving host got the packet (receive timestamp) and the time at which the receiving host sends the time it arrived. The difference between the arrival time and the original timestamp is the time required for the round trip. By calculating the difference between the receive and transmit timestamp, the host also can determine how long the other host took to respond once it got the request. By subtracting this from the round trip time the host can also estimate the transit time for both the request and the reply.

(ix) **Address Mask Request and Reply** – By using a few additional bits of the local ID part of an IP address, subnet numbers may be assigned to different physical networks. For example, a class B site can manage eight separate physical networks by using 3-bits from the local ID to specify the physical network. These bits together with the 16-bits IP address form the subnet number. To extract a subnet number from an IP address, an internal router uses a subnet address mask and perform a bitwise logical AND operation between it and the actual IP address. Next, do a bitwise logical AND between the mask and the IP address. The result is the 19-bit subnet number. A host can send an address mask request packet to a router to determine the address mask for the network to which it is attached. The router can respond with the reply.

(x) **Information Request and Reply** – These messages were originally designed to let a host determine its IP address when it starts up. Other protocols typically do this now and these two control messages have become obsolete.

Q.34. Explain the concept of subnet addressing. Illustrate your answer with the help of an example.

Ans. In *subnetting*, a network is divided into several smaller subnetworks with each subnetwork (or subnet). All hosts are now required to support subnet addressing. Instead of considering an IP address as just a network ID and host ID, the host ID portion is divided into a subnet ID and a host ID.

This makes sense because class A and class B addresses have too many bits allocated for the host ID— $2^{24}-2$ and $2^{16}-2$, respectively. People don't attach that many hosts to a single network. We subtract 2 in these expressions because host IDs of all zero bits or all one bits are invalid.

After obtaining an IP network ID of a certain class from the InterNIC, it is up to the local system administrator whether to subnet or not, and if so, how many bits to allocate to the subnet ID and host ID. For example, the Internet used in this text has a class B network address (140.252) and of the remaining 16 bits, 8 are for the subnet ID and 8 for the host ID. This is shown in fig. 1.30.

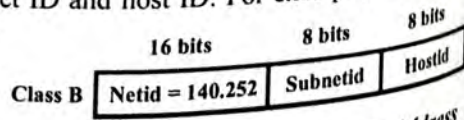


Fig. 1.30 Subnetting a Class B Address

This division allows 254 subnets, with 254 hosts per subnet.

Many administrators use the natural 8-bit boundary in the 16 bits of a class B host ID as the subnet boundary. This makes it easier to determine the subnet boundary for a class A or class B address, but there is no requirement that the subnet boundary for a class A or class B address be on a byte boundary.

Most examples of subnetting describe it using a class B address. Subnetting is also allowed for a class C address, but there are fewer bits to work with. Subnetting is rarely shown with a class A address because there are so few class A addresses.

Subnetting hides the details of internal network organization to external routers. Using our example network, all IP addresses have the class B network ID of 140.252. But there are more than 30 subnets and more than 400 hosts distributed over those subnets. A single router provides the connection to the Internet, as shown in fig. 1.31.

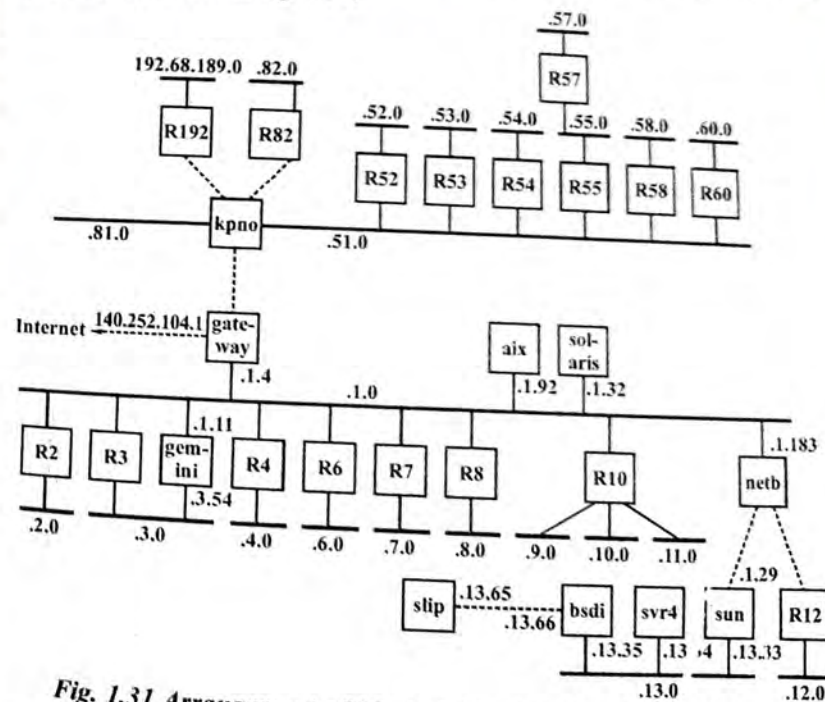


Fig. 1.31 Arrangement of Most of the noao.edu 140.252 Subnets

In this figure we have labeled most of the routers as R_n , where n is the subnet number. We show the routers that connect these subnets, along with the nine systems from the figure on the inside front cover. The Ethernets are shown as thicker lines and the point-to-point links as dashed lines. We do not show all the hosts on the various subnets. For example, there are more than 50 hosts on the 140.252.3 subnet and more than 100 on the 140.252.1 subnet.

Q.35. Describe the benefits of subnetting.

Ans. The benefits of subnetting are as follows –

- Smaller broadcast domains larger collision domains.
- Low-level security provided fewer broadcast domains.
- Increased address flexibility larger broadcast domains.

Q.36. Explain subnet mask.

Ans. Apart from the IP address, a host also wants to know how many bits are to be used for the host ID and how many bits are to be used for the subnet ID. A *subnet mask* specifies this at bootstrap time. This mask contains a 32-bit value having one bit for the network ID and subnet ID and zero bits for the host ID. The formation of the subnet mask for two different partitions of a class B address is shown in fig. 1.32.

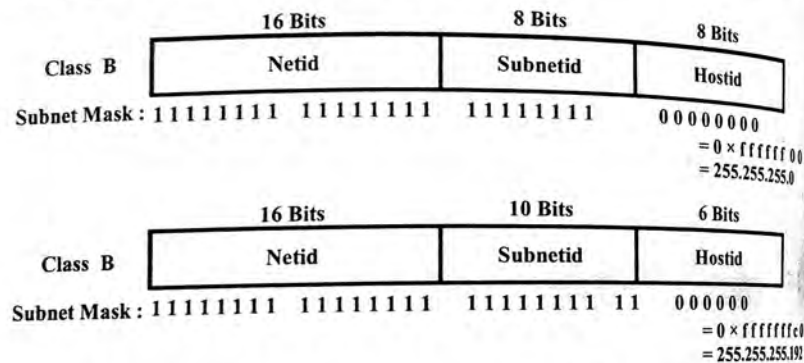


Fig. 1.32 Example Subnet Masks for Two Different Class B Subnet Arrangements

The top example shows a class B address partitioned for a 8 bit subnet ID and a 8 bit host ID. The lower example depicts class B address partitioned for a 10-bit subnet ID and a 6-bit host ID.

Also IP addresses are written in dotted-decimal notation, subnet masks are often written in hexadecimal, especially if the boundary is not a byte boundary, since the subnet mask is a bit mask.

Q.37. What is supernetting ?

Ans. An organization can combine several class C blocks to create a larger range of addresses, known as **supernetting**. In other words, several networks are combined to create a supernet. By doing this, an organization can apply for a set of class C blocks instead of just one. For example, an organization that needs 1000 addresses can be granted four class C blocks. The organization can then use these addresses in one supernet as shown in fig. 1.33.

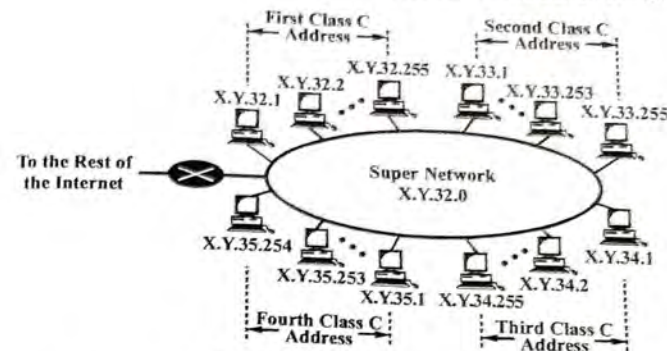


Fig. 1.33 A Supernetwork

Q.38. Discuss address assigning procedure of supernetting.

Ans. When we assign a set of class C blocks to an organization, we have two choices. We can choose the blocks randomly or based on a set of rules. If we choose the blocks randomly, the routers outside of the organization treat each block separately, they think each block belongs to a different site. In this way, each router will have N entries in its routing table where N is the number of blocks. This strategy may increase the size of the routing table tremendously.

The other choice is to make a superblock out of the blocks so that each router has only one entry in the routing table. To do so, we need to follow a set of rules when we assign the blocks –

- The number of blocks must be a power of 2.
- The blocks must be contiguous in the address space.
- The third byte of the first address in the superblock must be evenly divisible by the number of blocks. In other words, if the number of blocks is N, the third byte must be divisible by N.

Q.39. Explain supernet mask.

Ans. When an organization is granted one block of addresses (class A, B or C), the first address in the block and the mask define the block (range of addresses). We always know this range of addresses since the mask is always known (i.e. default mask).

When an organization divides its block into subnets, the first address in the subblock and the subnet mask completely define the subblock (i.e., the range of addresses). In this case, however, the first address alone is not enough, we must have the subnet mask.

When an organization combines several blocks into a superblock, we require to know the first address in the block and the supernet mask. The first address alone cannot define the range, we require a supernet mask to find how many blocks are combined to make a superblock.

A supernet mask is the reverse of a subnet mask. A subnet mask for class C has more 1s than the default mask for this class. A supernet mask for class C has less 1s than the default mask for this class.

Fig. 1.34 shows the difference between a subnet mask and a supernet mask. A subnet mask that divides a block into eight subblocks has three more 1s than the default mask. On the other hand, a supernet mask has three less 1s than the default eight blocks into one superblock has three less 1s than the default mask.

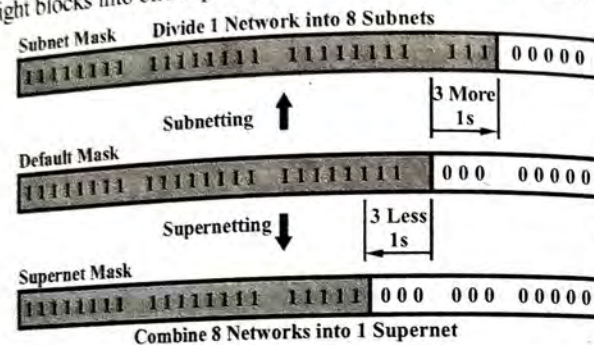


Fig. 1.34 Comparison of Subnet, Default and Supernet Masks

Q.40. Describe some special cases of IP addresses of subnetting.

Ans. There are seven special case IP addresses which are shown in table 1.4. In this, 0, means a field of all zero bits, -1 means a field of all one bits, and netid, hostid and subnetid means a field that is neither all zero bits nor all one bits. A blank subnet ID column means the address is not subnetted.

The table 1.4 is divided into three sections. The first two entries are special case source addresses, the next one is the special loopback address, and the final four are the broadcast addresses.

Table 1.4 Special Case IP Addresses

IP address			Can appear as		Description
Net ID	Subnet ID	Host ID	Source ?	Destination ?	
0		0	OK	Never	This host in this net (see restrictions below) specified host on this net (see restrictions below)
0		Hostid	OK	Never	
127		Anything	OK	OK	Loopback address
-1		-1	Never	OK	Limited broadcast (never forwarded)
Netid		-1	Never	OK	Net directed broadcast to netid
Netid		-1	Never	OK	Subnet directed broadcast to netid, subnet
Netid	Subnetid	-1	Never	OK	All subnets directed broadcast to netid

The first two entries in the table, with a network ID of 0, can only appear as the source address as part of an initialization procedure when a host is determining its own IP address, for example, when the BOOTP protocol is being used.

Q.41. Write short note on CIDR.

Ans. Some solutions solve one problem but create a new one. The solution that was implemented and which gave the Internet a bit of extra breathing room is CIDR. The fundamental idea behind CIDR, which is described in RFC 1519, is to allocate the remaining IP addresses in variable-sized blocks, without regard to the classes. When a site requires, say, 2000 addresses, it is given a block of 2048 addresses on a 2048-byte boundary.

Dropping the classes makes forwarding more complicated. In the old classful system, forwarding processed like this. If a packet arrived at a router, a copy of the IP address was shifted right 28 bits to provide a 4-bit class number. A 16-way branch then sorted packets into A, B, C, and D (if supported), with eight of the conditions for class A, four of the cases for class B, two of the conditions for class C, and one each for D and E. Then, the code for each class masked off the 8-, 16-, or 24-bit network number and right aligned it in a 32-bit word. The network number was then looked up in the A, B, or C table, generally by indexing for A and B networks and hashing for C networks. Once the entry was found, the out-going line could be looked up and the packet forwarded.

Q.42. What is DNS ?

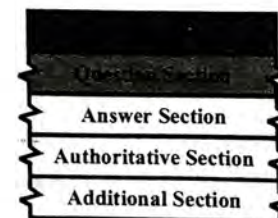
Ans. The DNS is a distributed database. It is used by TCP/IP applications to map between hostnames and IP addresses and tool provide electronic mail routing information. We use the term distributed because no single site on the Internet knows all the information. Each site maintains its own database of information and runs a server program that other system across the Internet (clients) can query. The DNS provides the protocols that allow client and server to communicate with each other.

Q.43. Explain message format of DNS.

Ans. DNS has two types of messages namely, *Query* and *response*. Both types have the same format. The Query message consists of a header and the question records, whereas, the response message consists of a header, question records, answer records, authoritative records and additional records as shown in fig. 1.35.



(a) Query



(b) Response

Fig. 1.35 Query and Response Messages

(i) **Header** – Both query and response messages have the same header format with some fields set to zero for the query messages. The header is 12 bytes and its format is shown in fig. 1.36.

Identification	Flags
Number of Question Records	Number of Answer Records (All 0s in query message)
Number of Authoritative Records (All 0s in query message)	Number of Additional Records (All 0s in query message)

Fig. 1.36 Header Fields

Fig. 1.36 Header Format

The header fields are as follows –

(a) **Identification** – This is a 16-bit field used by the client to match the response with the query. The client uses a different identification number each time it sends a query. The server duplicates this number in the corresponding response.

(b) **Flags** – This is a 16-bit field consisting of the subfields shown in fig. 1.37.

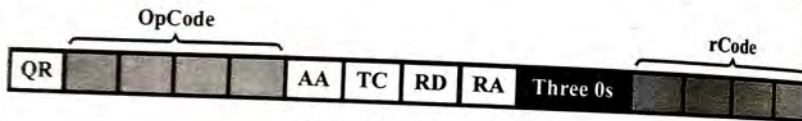


Fig. 1.37 Flags Field

A brief description of each flag subfield is as follows –

(1) **QR (Query/Response)** – This is a 1-bit subfield that defines the type of message. If it is 0, the message is a query. If it is 1, the message is response.

(2) **OpCode** – This is a 4-bit subfield that defines the type of query or response (0 if standard, 1 if inverse and 2 if a server status request).

(3) **AA (Authoritative Answer)** – This is a 1-bit subfield. When it is set (value of 1) it means that the name server is an authoritative server. It is used only in a response message.

(4) **TC (Truncated)** – This is a 1-bit subfield. When it is set (value of 1), it means that the response was more than 512 bytes and truncated to 512. It is used when DNS uses the services of UDP.

(5) **RD (Recursion Desired)** – This is a 1-bit subfield. When it is set (value of 1) it means the client desires a recursive answer. It is set in the query message and repeated in the response message.

(6) **RA (Recursion Available)** – This is a 1-bit sub-field. When it is set in the response, it means that a recursive response is available. It is set only in the response message.

(7) **Reserved** – This is a 3-bit subfield set to 000.

(8) **rCode** – This is a 4-bit field that shows the status of the error in the response. Of course, only an authoritative server can make such a judgement. Table 1.5 shows the possible values for this field.

Table 1.5 Values of rCode

Value	Meaning
0	No error
1	Format error
2	Problem at name server
3	Domain reference problem
4	Query type not supported
5	Administratively prohibited
6-15	Reserved

(c) **Number of Question Records** – This is a 16-bit field containing the number of queries in the question section of the message.

(d) **Number of Answer Records** – This is a 16-bit field containing the number of answer records in the answer section of the response message. Its value is zero in the query message.

(e) **Number of Authoritative Records** – This is a 16-bit field containing the number of authoritative records in the authoritative section of a response message. Its value is zero in the query message.

(f) **Number of Additional Records** – This is a 16-bit field containing the number of additional records in the additional section of a response message. Its value is zero in the query message.

(ii) **Question Section** – This is a section consisting of one or more question records. It is present on both query and response messages.

(iii) **Answer Section** – This is a section consisting of one or more resource records. It is present only on response messages. This section includes the answer from the server to the client (resolver).

(iv) **Authoritative Section** – This is a section consisting of one or more resource records. It is present only on response messages. This section gives information (domain name) about one or more authoritative servers for the query.

(v) **Additional Information Section** – This is a section consisting of one or more resource records. It is present only on response messages. This section provides additional information that may help the resolver.

Q.44. Discuss about the domain name space.

Ans. A **domain name space** was designed to have a hierarchical name space. In this design the names are defined in an inverted-tree structure with the root at the top. The tree can have only 128 levels ranging from level 0 (root) to level 127.

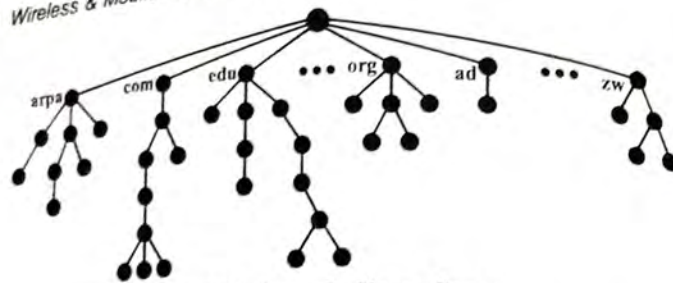


Fig. 1.38 Domain Name Space

Each node in the tree has a **label**, which is a string with a maximum of 63 characters. The root label is a special case that contains a null string. DNS requires that children of a node have different labels, which guarantees the uniqueness of the domain names.

Each node in the tree has a domain name. A full **domain name** is a sequence of labels separated by dots (.). The domain names are always read from the node up to the root. The last label is the label of the root i.e., a null. It means that a full domain name always ends in a null label, which means the last character is a dot because the null string is nothing. Fig. 1.39 shows some domain names.

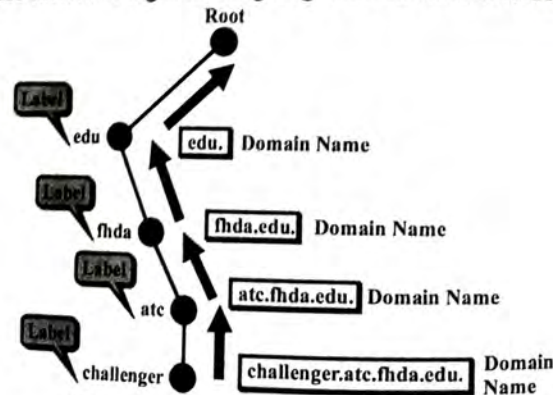


Fig. 1.39 Domain Names and Labels

If a label is terminated by a null string, it is known as a fully qualified domain name (FQDN). A FQDN is a domain name that contains the full name of a host. It contains all labels, from the most specific to most general, that uniquely defines the name of the host. For instance, the domain name `challenger.atc.fhda.edu.` is the FQDN of a computer named challenger and installed at the Advanced Technology Center (ATC) at De Anza College. A DNS server can only match an FQDN to an address.

If a label is not terminated by a null string, it is called a partially qualified domain name (PQDN). A PQDN begins from a node but it does not reach the root. It is used when the name to be resolved belongs to the same site as the

client. Here the resolver can supply the missing part, called the suffix, to create an FQDN. For instance, if a user at the `fhda.edu.` site wants to get the IP address of the challenger computer, he or she can define the partial name *challenger*. The DNS client adds the suffix `atc.fhda.edu.` before passing the address to the DNS server.

A subtree of domain name space is called **domain**. The name of the domain is the domain name of the node at the top of the subtree.

Q.45. Discuss DNS in the Internet.

Ans. The domain name space (tree) is divided in three different sections in the Internet –

(i) **Generic Domains** – The generic domains define registered hosts according to their generic behaviour. Each node in the tree defines a domain, which is an index to the domain name space database, as shown in fig. 1.41.

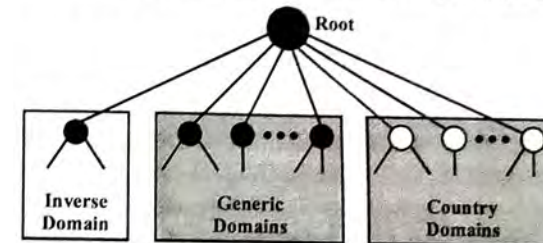


Fig. 1.40 DNS-used in the Internet

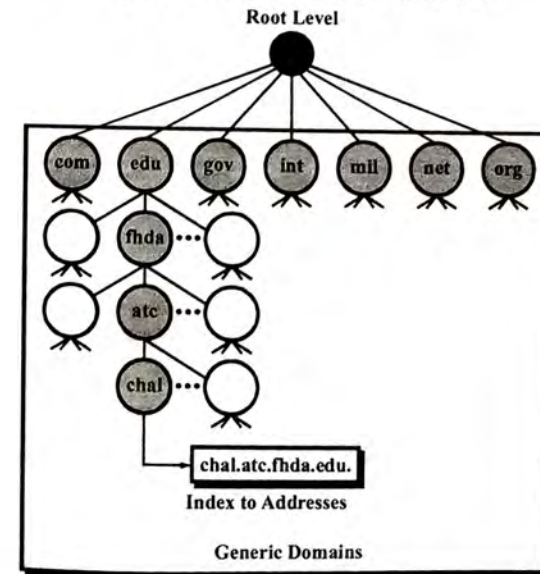


Fig. 1.41 Generic Domains

Looking at the tree, we see that the first level in the generic domains section allows seven possible three-character labels. These labels describe the organization types as shown in table 1.6. Recently a few more first-level labels has been approved, these are shown in table 1.7.

Table 1.6 Generic Domain Labels

Label	Description
com	Commercial organizations
edu	Educational institutions
gov	Government institutions
int	International organizations
mil	Military groups
net	Network support centers
org	Nonprofit organizations

Table 1.7 New Generic Domain Labels

Label	Description
aero	Airlines and aerospace companies
biz	Businesses or firms (similar to "com")
coop	Cooperative business organizations
info	Information service providers
museum	Museums and other non-profit organizations
name	Personal names (individuals)
pro	Professional individual organizations

(ii) **Country Domains** – The country domains section follows the same format as the generic domains but uses two-character country abbreviations (e.g., us for United States) in place of the three-character organizational abbreviations at the first level. Second-level labels can be organizational or they can be more specific, national designations. The United States, for example, uses state abbreviations as a subdivision of us (e.g., ca.us.).

Fig. 1.42 shows the country domains section. The address *anza.cup.ca.us* can be translated to De Anza College in Cupertino in California in the United States.

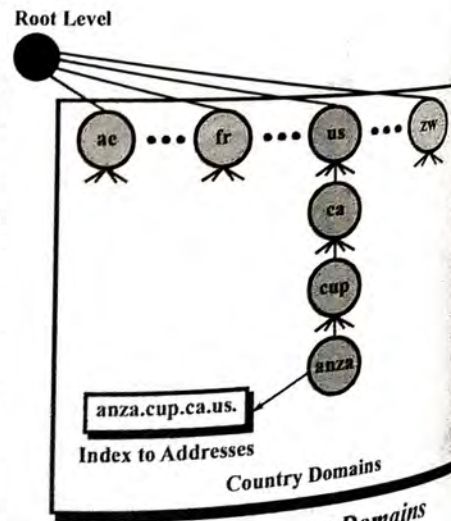


Fig. 1.42 Country Domains

(iii) **Inverse Domains** – The inverse domain is used to map an address to a name. This may happen, for example, when a server has received a request from a client to do a task. Whereas the server has a file that contains a list of authorized clients, the server lists only the IP address of the client. To determine if the client is on the authorized list, it can ask its resolver to send a query to the DNS server and ask for a mapping of address to name.

This type of query is called an inverse or pointer (PTR) query. To handle a pointer query, the inverse domain is added to the domain name space with the first-level node called *arpa* (for historical reasons). The second level is also one single node named *in-addr* (for inverse address). The rest of the domain defines IP addresses.

The servers that handle the inverse domain are also hierarchical. It means the netid part of the address should be at a higher level than the subnetid part and the subnetid part higher than the hostid part. In this way, a server serving the whole site is at a higher level than the servers serving each subnet. This configuration makes the domain look inverted when compared to a generic or country domain. To follow the convention of reading the domain labels from the bottom to the top, an IP address such as 132.34.45.121 is read as 121.45.34.132.in-addr.arpa, as shown in fig. 1.43.

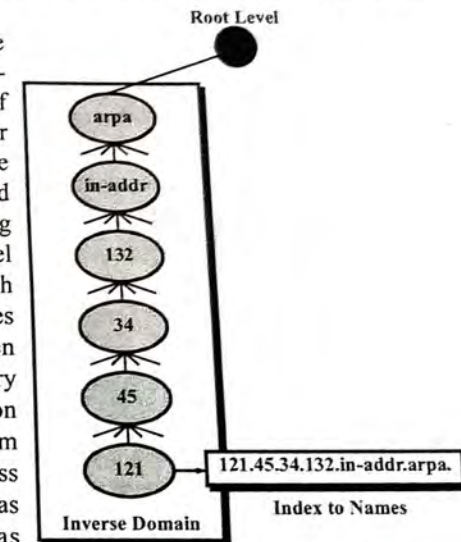


Fig. 1.43 Inverse Domain

NAT – PRIVATE ADDRESSING AND NAT, SNAT, DNAT, NAT AND FIREWALLS

Q.46. What do you mean by network address translation (NAT) ?

Ans. To access the Internet, one public IP address is needed but we can use a private IP address in our private network. The idea of NAT is to allow multiple devices to access the Internet through a single public address. To achieve this, the translation of private IP address to a public IP address is required. Network address translation (NAT) is a process in which one or more local IP address is translated into one or more Global IP address and vice versa in order to provide Internet access to the local hosts. Also, it does the

translation of port numbers i.e. masks the port number of the host with another port number, in the packet that will be routed to the destination. It then makes the corresponding entries of IP address and port number in the NAT table. NAT generally operates on router or firewall.

Generally, the border router is configured for NAT i.e., the router which has one interface in local (inside) network and one interface in the global (outside) network. When a packet traverse outside the local (inside) network, then NAT converts that local (private) IP address to a global (public) IP address. When a packet enters the local network, the global (public) IP address is converted to a local (private) IP address.

If NAT run out of addresses i.e., no address is left in the pool configured then the packets will be dropped and an Internet control message protocol (ICMP) host unreachable packet to the destination is sent.

Q.47. Write down the advantages and disadvantages of NAT.

Ans. Advantages of NAT – Advantages of NAT are given below –

- (i) NAT conserves legally registered IP addresses.
- (ii) It provides privacy as the device IP address, sending and receiving the traffic, will be hidden.
- (iii) Eliminates address renumbering when a network evolves.

Disadvantages of NAT – Disadvantages of NAT are given below –

- (i) Translation results in switching path delays.
- (ii) Certain applications will not function while NAT is enabled.
- (iii) Complicates tunneling protocols such as IPsec.
- (iv) Also, router being a network layer device, should not tamper with port numbers (transport layer) but it has to do so because of NAT.

Q.48. Discuss the ways to configure the network address translation.

Ans. There are three ways to configure NAT –

- (i) **Static NAT** – In this, a single unregistered (private) IP address is mapped with a legally registered (public) IP address i.e., one-to-one mapping between local and global address. This is generally used for web hosting. These are not used in organizations as there are many devices who will need Internet access and to provide Internet access, the public IP address is needed.
- (ii) **Dynamic NAT** – In this type of NAT, an unregistered IP address is translated into a registered (public) IP address from a pool of public IP address. If the IP address of pool is not free, then the packet will be dropped as an only a fixed number of private IP address can be translated to public addresses.

- (iii) **Port Address Translation (PAT)** – This is also known as NAT overload. In this, many local (private) IP addresses can be translated to a single registered IP address. Port numbers are used to distinguish the traffic

i.e., which traffic belongs to which IP address. This is most frequently used as it is cost-effective as thousands of users can be connected to the Internet by using only one real global (public) IP address.

Q.49. Define the following terms –

- (i) NAT inside and outside addresses
- (ii) Inside local address
- (iii) Inside global address
- (iv) Outside local address
- (v) Outside global address.

Ans. (i) NAT Inside and Outside Addresses – Inside refers to the addresses which must be translated. Outside refers to the addresses which are not in control of an organization. These are the network addresses in which the translation of the addresses will be done.

(ii) Inside Local Address – An IP address that is assigned to a host on the inside (local) network. The address is probably not a IP address assigned by the service provider i.e., these are private IP address. This is the inside host seen from the inside network.

(iii) Inside Global Address – IP address that represents one or more inside local IP addresses to the outside world. This is the inside host as seen from the outside network.

(iv) Outside Local Address – This is the actual IP address of the destination host in the local network after translation.

(v) Outside Global Address – This is the outside host as seen from the outside network. It is the IP address of the outside destination host before translation.

Q.50. What is firewall ? Explain the design principle of firewall.

(R.G.P.V., June 2011, Dec. 2013, 2016)

Or

List three design goals of a firewall.

(R.G.P.V., Dec. 2010)

Or

Write and explain firewall design principles.

(R.G.P.V., Dec. 2017)

Or

Write short note on firewall design principle.

(R.G.P.V., Nov. 2018)

Ans. A **firewall** defines a single choke point that keeps unauthorized users out of the protected network, prohibits potentially vulnerable services from entering or leaving the network, and provides protection from various kinds of IP spoofing and routing attacks. The use of a single choke point simplifies security management because security capabilities are consolidated on a single system or set of systems.

Following are the design goals or design principles of firewall –

- (i) All traffic from inside to outside, and vice versa, must pass through the firewall. This is achieved by physical blocking all access to the local network except via the firewall.
- (ii) Only authorized traffic will be allowed to pass. Various types of firewalls are used, which implement various types of security policies.
- (iii) The firewall itself is immune to penetration. This implies that use of a trusted system with a secure operating system.

Q.51. Explain the firewall techniques.

Ans. The four general techniques, that are used by firewalls to control access and enforce the site's security policy, are as follows –

(i) **Service Control** – Determines the types of Internet services that can be accessed, inbound or outbound. The firewall may filter traffic on the basis of IP address and TCP port number. It may provide proxy software that receives and interprets each service request before passing it on. It may also host the server software itself, such as a Web or mail service.

(ii) **Direction Control** – Determines the direction in which particular service requests may be initiated and allowed to flow through the firewall.

(iii) **User Control** – Controls access to a service according to which user is attempting to access it. This feature is typically applied to users inside the firewall perimeter (local users). It may also be applied to incoming traffic from external users. The latter requires some form of secure authentication technology, such as is provided in IPSec.

(iv) **Behaviour Control** – Controls how particular services are used. For example, the firewall may filter e-mail to eliminate spam, or it may enable external access to only a portion of the information on a local Web server.

Q.52. List various advantages and disadvantages of firewalls.

Ans. The following capabilities are within the scope of a firewall –

- (i) A firewall provides a location for monitoring security-related events. Audits and alarms can be implemented on the firewall system.
- (ii) A firewall can serve as the platform for IPSec. Using the tunnel mode capability, the firewall can be used to implement virtual private networks.
- (iii) A firewall is a convenient platform for several Internet functions that are not security related. These include a network address translator, which maps local addresses to Internet addresses, and a network management function that audits or logs Internet usage.

Firewalls have the following disadvantages –

- (i) The firewall does not protect against internal threats, such as a disgruntled employee or an employee who unwillingly cooperates with an external attacker.

(ii) The firewall cannot protect against the transfer of virus-infected programs or files. Because of the variety of operating systems and applications supported inside the perimeter, it would be impractical and perhaps impossible for the firewall to scan all incoming files, e-mail, and messages for viruses.

(iii) The firewall cannot protect against attacks that bypass the firewall. Internal systems may have dial-out capability to connect to an ISP. An internal LAN may support a modem pool that provides dial-in capability for travelling employees and telecommuters.

Q.53. What is firewall design principle ? How it work against attacks and threats.
(R.G.P.V., Dec. 2012)

Ans. Firewall Design Principle – Refer the ans. of Q.50.

Working of Firewall Against Attacks and Threats – A firewall can work in different manner to protect the network such as a filtering device, a border sentry, a gateway and a traffic control device.

Being a filtering device, it allows what you want on your network and prohibits what you do not. Filtering relies on filtering rules. Each rule has a pattern of concern and a response the firewall will make if an incoming element matches the pattern. Firewalls follow a philosophy of security known as deny by default, allow by exception. Firewall filtering compares each packet received to a set of exception rules. These rules state that content in the packet is either. If the packet matches a deny rule, then the packet is dropped. If the packet matches an allow rule, it continues on to its destination. If a packet fails to match any rule then the firewall drops the packet by default.

Network security administrators use the term firewall to refer to a device that allows some traffic but denies other traffic. In network security, firewalls allows traffic through that's considered safe – or at least authorizes traffic without the entire firewall failing.

A gateway is an entrance or exit point to access a controlled space. Gateways are important because high levels of traffic pass through a gateway. Thus, positioning a firewall at a network gateway is an aspect of secure design.

A firewall stands as a sentry against all attacks and attempts at system compromise. A firewall protects the network against the substantial asset damage that external attacks can cause.

Q.54. Explain different types of firewall and discuss design principle of firewall.
(R.G.P.V., Dec. 2014)

Ans. Fig. 1.44 shows the three common types of firewalls – packet-filters, application-level gateways, and circuit-level gateways.

(i) **Packet-filtering Router** – A packet-filtering router shown in Fig. 1.44(a), applies a set of rules to each incoming IP packet and then forwards or discards the packet. The router is typically configured to filter packets

going in both directions (from and to the internal network) based on information contained in a network packet – source IP address, destination IP address, source and destination transport-level address, protocol field and the interface.

The packet filter is typically setup as a list of rules based on matching fields in the IP or TCP header. If there is a match to one of the rules, that is invoked to determine whether to forward or discard the packet. If there is no match to any rule, then a default action is taken. Two default policies are possible –

(a) **Default = discard** – That which is not expressly permitted is prohibited.

(b) **Default = forward** – That which is not expressly prohibited is permitted.

The default discard scheme is the more conservative. Initially, everything is blocked, and services must be added on a case-by-case basis. This policy is more visible to users, who are more likely to see the firewall as a hindrance. The default forward scheme increases ease of use for end users but provides reduced security. In essence, the security administrator must react to a new security threat as it becomes known.

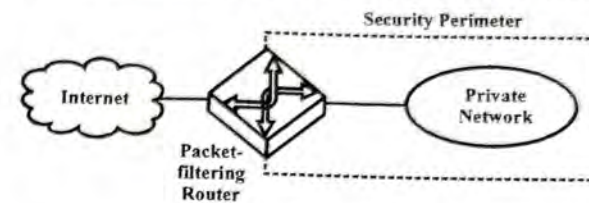
One advantage of a packet-filtering router is its simplicity. Also, packet filters are transparent to users and are very fast.

(ii) **Application-level Gateway** – An application-level gateway, known as a *proxy server*, acts as a relay of application-level traffic as shown in fig. 1.44 (b). The user contacts the gateway using a TCP/IP application, such as Telnet or FTP, and the gateway asks the user for the name of the remote host to be accessed. When the user responds and inputs a valid user ID and authentication information, the gateway contacts the application on the remote host and relays TCP segments containing the application data between the endpoints. If the gateway does not implement the proxy code for a specific application, the service is not supported and cannot be forwarded across the firewall. Further, the gateway can be configured to support only specific features of an application that the network administrator considers acceptable while denying all other features.

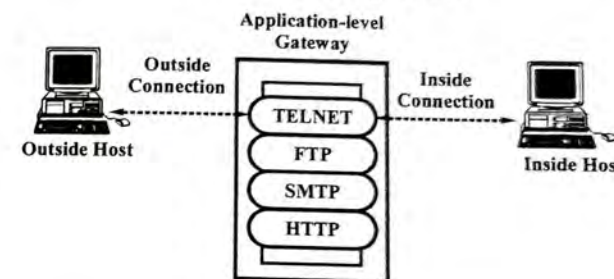
Application-level gateways tend to be more secure than packet-filtering gateways. Rather than trying to deal with the numerous possible combinations that can be allowed and forbidden at the TCP and IP level, the application level gateway need only scrutinize a few allowable applications. In addition, it is easy to audit and audit all incoming traffic at the application level.

A prime disadvantage of the application-level gateway is the additional processing overhead on each connection. In effect, there are two separate

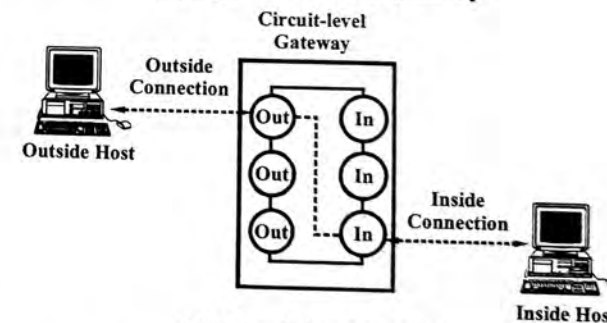
connections between the end users, with the gateway at the splice point, and the gateway must examine and forward all traffic in both directions.



(a) Packet-filtering Router



(b) Application-level Gateway



(c) Circuit-level Gateway

Fig. 1.44 Three Types of Firewalls

(iii) **Circuit-level Gateway** – A third type of the firewall is the circuit-level gateway shown in fig. 1.44 (c). It can be a stand-alone system or it can perform a specialized function performed by an application-level gateway for certain applications. A circuit-level gateway does not allow an end-to-end TCP connection. Rather, the gateway sets up two TCP connections, one between itself and a TCP user on an inner host and other between itself and a TCP user on an outside host. Once the two connections are established, the gateway typically relays TCP segments from one connection to other without examining

that were present in the previous configuration are present here as well. An information server or other hosts can be allowed direct communication with the router if this is in accord with the security scheme.

(iii) Screened Subnet Firewall System – This configuration is shown in fig. 1.45 (c). This is the most secure configuration. In this configuration two packet-filtering routers are used, one between the bastion host and the Internet and the other between the bastion host and the internal network. This configuration creates an isolated subnetwork, which may consist of servers, modems for dial-in capability. Typically, both the Internet and the internal network have access to hosts on the screened subnet, but traffic across the screened subnet is blocked. This configuration has following advantages

- There are now three levels of defense to thwart intruders.
- The outside router advertises only the existence of the screened subnet to the Internet. Therefore, the internal network is invisible to the Internet.
- Similarly, the inside router advertises only the existence of the screened subnet to the internal network. Therefore, the systems on the inside network cannot construct direct routes to the Internet.

VLANS – CONCEPTS, COMPARISON WITH REAL LANS, TYPES OF VLAN, TAGGING

Q.57. What is the concept of virtual LAN (VLAN) ?

Ans. The concept of VLAN is to divide a LAN into logical, instead of physical, segments. A LAN can be divided into many logical LANs called VLANs. Fig. 1.46 shows the switched LAN divided into VLANs. Each VLAN

Switch with VLAN Software

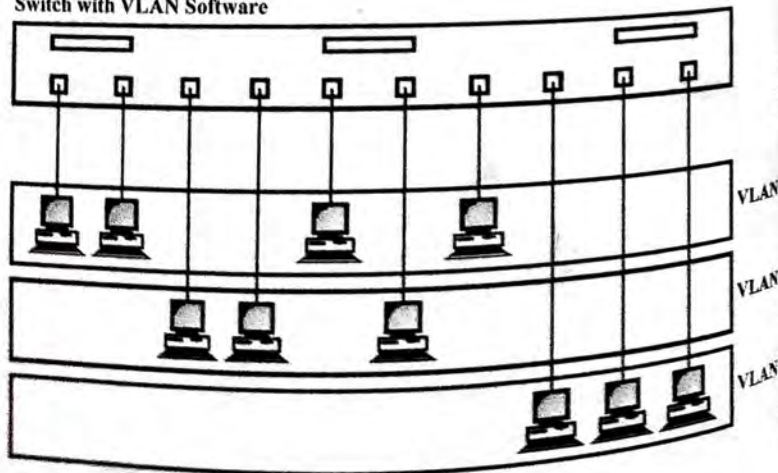


Fig. 1.46

a work group in the organization. If an engineer moves from one group to another, there is no need to change the physical configuration. Any station can be logically moved to another virtual LAN. In virtual LANs, the group membership is defined by software, not hardware. All members belonging to a VLAN can receive broadcast messages sent to that particular VLAN.

If a station moves from VLAN 1 to VLAN 2, it receives broadcast messages sent to VLAN 2, but no longer receives broadcast messages sent to VLAN 1.

Virtual LAN technology even permits the grouping of stations connected to different switches in a virtual LAN. Fig. 1.47 shows a backbone local area network with three VLANs and two switches. Stations from switches A and B belong to each VLAN.

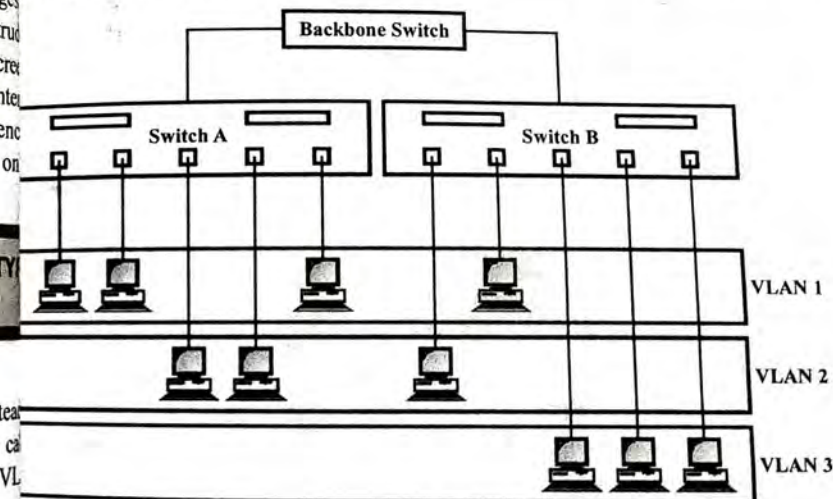


Fig. 1.47

In a company, this is a good configuration with two separate buildings. An engineer in the first building and engineer in the second building can be in the same work group even though they are connected to different physical LANs. Each building can have its own switched LAN connected by a backbone.

Q.58. Give the advantages of virtual LAN.

Ans. Advantages of virtual LAN are as follows –

- Performance** – The network traffic is full of broadcast and multicast. VLAN reduces the need to send such traffic to unnecessary destinations. e.g. if the traffic is intended for 2 users but as 10 devices are present in the same broadcast domain therefore all will receive the traffic i.e., wastage of bandwidth but if we make VLANs, then the broadcast or multicast packet will go to the intended users only.

(ii) **Security** – In the same network, sensitive data can be broadcast domains, set up firewalls, restrict access. Also, VLANs can be used to inform the network manager of an intrusion. Hence, VLANs greatly enhance network security.

(iii) **Formation of Virtual Groups** – As there are different departments in every organization namely sales, finance etc., VLANs can be very useful in order to group the devices logically according to their departments.

(iv) **Cost Reduction** – VLANs can be used to create broadcast domains which eliminate the need for expensive routers.

(v) **Flexibility** – VLAN provide flexibility to add, remove the members of host we want.

(vi) By using VLAN, the number of small size broadcast domains can be increased which are easy to handle as compared to a bigger broadcast domain.

Q.59. Give the comparison between LAN and VLAN.

Ans. Comparison between LAN and VLAN are as follows –

S.No.	LAN	VLAN
(i)	LAN stands for local area network.	VLAN stands for virtual local area network.
(ii)	The cost of local area network is high.	The cost of virtual local area network is less.
(iii)	The latency of local area network is high.	The latency of virtual local area network is low.
(iv)	The devices which are used in LAN are – hubs, routers and switch.	The devices which are used in VLAN are – bridges and switches.
(v)	In local area network, the packet is advertised to each device.	In virtual local area network, packet is send to specific broadcast domain.
(vi)	Local area network is less efficient than virtual local area network.	Virtual local area network is more efficient than local area network.
(vii)	Local area network work on single broadcast domain.	Virtual local area network work on multiple broadcast domains.

Q.60. Discuss the various types of virtual LANs depending on network type.

Ans. Types of VLANs depending on the type of the network are as follows:

(i) **Default VLAN** – When the switch initially starts up, all ports become a member of the default VLAN (generally all switches

have a default VLAN named as VLAN 1), which makes them all part of the same broadcast domain. Using default VLAN allows any network device connected to any of the switch port to connect with other devices on other switch ports. A unique feature of default VLAN is that it cannot be rename or delete.

(ii) **Voice VLAN** – Voice VLAN is configured to carry voice traffic. Voice VLANs are mostly given high transmission priority over other types of network traffic. To ensure voice over IP (VoIP) quality (delay of less than 150 milliseconds (ms) across the network), we must have separate voice VLAN. This will preserve bandwidth for other applications.

(iii) **Data VLAN** – Data VLAN is used to divide the whole network into 2 groups. One group of users and other group of devices. This VLAN is known as a user VLAN, the data VLAN is used only for user-generated data. This VLAN carrying data only. It is not used for carrying management traffic or voice.

(iv) **Native VLAN** – This VLAN identifies traffic coming from each end of a trunk link. A native VLAN is allocated only to an 802.1Q trunk port. The 802.1Q trunk port places untagged traffic (traffic that does not come from any VLAN) on the native VLAN. It is a best to configure the native VLAN as an unused VLAN.

(v) **Management VLAN** – A management VLAN is configured to access the management capabilities of a switch (traffic like system logging, monitoring). VLAN 1 is the management VLAN by default (VLAN 1 would be a bad choice for the management VLAN). Any of a switch VLAN could be configured as the management VLAN if admin as not configured a unique VLAN to serve as the management VLAN. This VLAN ensures that bandwidth for management will be available even when user traffic is high.

Q.61. Write short note on virtual LAN tagging.

Ans. VLAN tagging is a method through which more than one VLAN is bundled on a port. VLAN tagging is used to tell which packet belongs to which VLAN on the other side. To make recognition easier, a packet is tagged with a VLAN tag in the Ethernet frame. Independent logical systems can be formed accurately with the help of the VLAN tagging inside a physical network itself. Individual domains can be created with the help of this VLAN tagging system.

VLAN tagging is performed by the putting the VLAN ID into a header to identify which network it is present in. This helps in determining which interface in a broadcast area the information packet needs to be sent to in order to receive the right information. The switches need to be configured beforehand for working properly with the process of VLAN tagging. With this system, multiple broadcast systems can be segregated into individual domains. Bridging traffic can be forwarded with the use of this system clients and information can be organized, configured and grouped.

Q.62. Define IPv6.

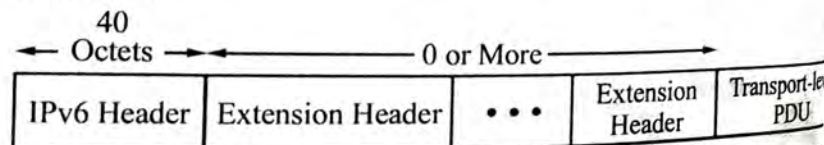
Ans. IPv6, also known as IPng (IP next generation), uses 128-bit addresses, versus the 32-bit (4-byte) addresses currently used in version 4. IPv6 can nearly accommodate a larger number of users. In version 6, the packet format has been simplified, yet at the same time it is more flexible. The new version supports authentication, data integrity, and confidentiality at the network layer. It is designed to handle the transmission of real-time data such as audio and video, and can carry data from other protocols. IPng can also handle congestion and route discovery better than the current version.

Q.63. The protocol field used in the IPv4 header is not present in the fixed IPv6 header. Why not?

Ans. The protocol field tells the destination host which protocol it should give the IP packet to. Intermediate routers do not need this information if it is not needed in the main header. Actually, it is there, but disguised. The Next Header field of the last (extension) header is used for this purpose.

Q.64. Discuss the IPv6 structure in brief.

Ans. An IPv6 protocol data unit (known as a packet) has the following general form –



The only header that is referred to simply as the IPv6 header. This is of fixed size with a length of 40 octets compared to 20 octets for the mandatory portion of the IPv4 header. The following extension headers have been defined –

- (i) **Hop-by-Hop Options Header** – Defines special options that require hop-by-hop processing.
- (ii) **Routing Header** – Provides extended routing, similar to IPv4 source routing.
- (iii) **Fragment Header** – Contains fragmentation and reassembly information.
- (iv) **Authentication Header** – Provides packet integrity and authentication.
- (v) **Encapsulating Security Payload Header** – Provides privacy.

(vi) **Destination Options Header** – Contains optional information to be examined by the destination node.

The IPv6 standard recommends that, when multiple extension headers are used, the IPv6 headers appear in the following order.

- (i) IPv6 header – Mandatory must always appear first
- (ii) Hop-by-Hop options header
- (iii) Destination options header – For options to be processed by the first destination that appears in the IPv6 Destination Address field plus subsequent destinations listed in the routing header
- (iv) Routing header
- (v) Fragment header
- (vi) Authentication header
- (vii) Encapsulating security payload header
- (viii) Destination option header for options to be processed only by the final destination of the packet.

Fig. 1.48 shows an example of an IPv6 packet that includes an instance of each header, except those related to security.

Note that the IPv6 header and each extension header include a Next Header field. This field identifies the type of the immediately following header. If the next header is an extension header, then this field contains the type identifier of that header. Otherwise, this field contains the protocol identifier of the upper-layer protocol using IPv6 (typically a transport-level protocol), using the same values as the IPv4 protocol field. In fig. 1.48, the upper-layer protocol is TCP, so that the upper-layer data carried by the IPv6 packet consist of a TCP header followed by a block of application data.

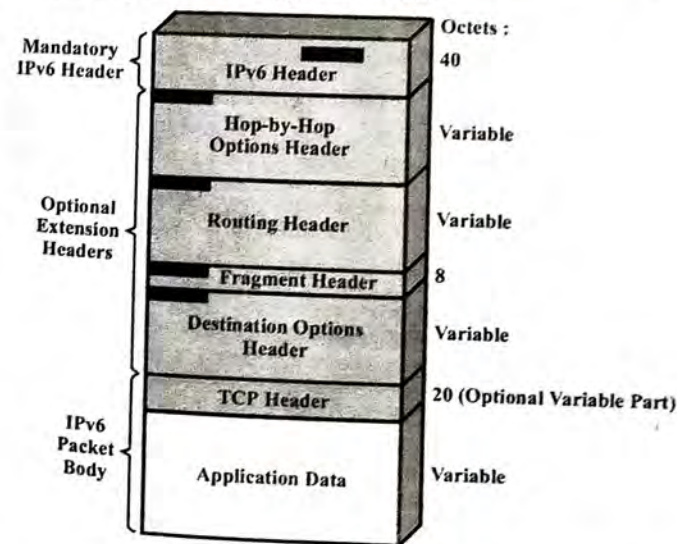


Fig. 1.48 IPv6 Packet with Extension Headers

Q.65. How is address space assignment done in IPv6?

- Ans. The partition of IPv6 address space has generated two issues:
- How to manage address assignment
 - How to map an address to a route.

The first issue focuses on the problem of devising a hierarchy of authority. The large address space in IPv6 permits a multilevel hierarchy or the hierarchies. The second issue focuses on computational efficiency. Independent of the hierarchy of authority that assigns addresses, a router must examine each datagram and choose a path to the destination. To keep the cost of speed routers low, the processing time required to choose a path must be kept low.

Although the first 8 bits of an address are sufficient to identify its address space, the address space is not partitioned into sections of equal size. (see table)

Table 1.8 The Proposed IPv6 Address Space Assignment

Binary Prefix	Type of Address	Part of Address Space
0000 0000	Reserved (IPv4 compatibility)	1/256
0000 0001	Unassigned	1/256
0000 001	NSAP addresses	1/128
0000 010	IPX addresses	1/128
0000 011	Unassigned	1/128
0000 1	Unassigned	1/32
0001	Unassigned	1/16
001	Aggregatable global unicast	1/8
010	Unassigned	1/8
011	Unassigned	1/8
100	Unassigned	1/8
101	Unassigned	1/8
110	Unassigned	1/16
1110	Unassigned	1/32
1111 0	Unassigned	1/64
1111 10	Unassigned	1/128
1111 110	Unassigned	1/512
1111 1110 0	Unassigned	1/1024
1111 1110 10	Link-local unicast addresses	1/1024
1111 1110 11	Site-local unicast addresses	1/256
1111 1111	Multicast addresses	

Only 15% of the address space has been assigned at present. Despite the sparse assignment, addresses have been chosen to make processing more efficient.

Q.66. Explain the following two types of headers contained in IPv6:

- IPv6 header (base header)
- Extension header

Ans. (i) **IPv6 Header** – The IPv6 header has a fixed length of 40 octets, consisting of the following fields (fig. 1.49) –

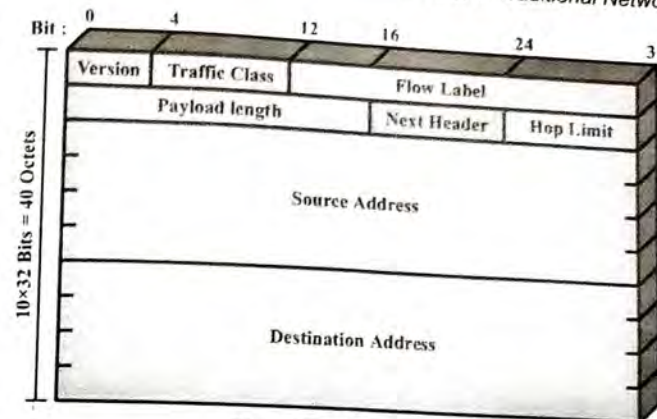


Fig. 1.49 IPv6 Header

(a) **Version (4 bits)** – Internet protocol version number; the value is 6.

(b) **Traffic Class (8 bits)** – Available for use by originating nodes and/or forwarding routers to identify and distinguish between different classes or priorities of IPv6 packets. This field is now used for the 0s and 1s fields, as for the IPv4 TOS field.

(c) **Flow Label (20 bits)** – May be used by a host to label those packets for which it is requesting special handling by routers within a network.

(d) **Payload Length (16 bits)** – This is the total length of all of the extension headers plus the transport level PDU.

(e) **Next Header (8 bits)** – Identifies the type of header immediately following the IPv6 header; this will either be an IPv6 extension header or a higher-layer protocol header, such as TCP or UDP.

(f) **Hop Limit (8 bits)** – The hop limit is an eight-bit field that serves the same purpose as the time-to-live (TTL) field used with IPv4. The hop limit field holds an eight-bit unsigned integer that is decremented by 1 by each router that forwards the packet.

(g) **Source Address (128 bits)** – This field carries the 128-bit address of the device that originates the packet.

(h) **Destination Address (128 bits)** – The destination address is the only options header that can appear in more than one place. It is a 128-bit field that carries the 128-bit IP address of the device intended recipient of the packet, which may not be the ultimate destination if an optional routing header is used.

(ii) **Extension Header** – There are currently six extension headers available with IPv6, discussed below –

(a) **Hop-by-Hop Options Header** – It allows the source of IP packet to transfer information to all the routes visited by the packet, as management, debugging, and certain control functions. So far, only three options have been defined – **Pad1**, **PadN** and **jumbo payload**. The **Pad1** option is 1 byte long and is designed for alignment purposes. **PadN** is similar concept to **Pad1**. The difference is that **PadN** is used when 2 or more bytes are needed for alignment. The **jumbo payload** option is used to define a payload longer than 65,535 bytes.

(b) **Source Routing Header** – The source routing option is essentially a combination of two IPv4 options – the strict source route option and the loose source route option.

(c) **Fragment Header** – With IPv6, the concept of fragmentation is identical to IPv4 in the sense that the length of a packet cannot exceed the MTU (maximum transfer unit) of the network through which it is being transported. In IPv6, only the original source can fragment. A source must use a **path MTU discovery technique** to find the smallest MTU supported by the network on the path. The source then fragments using this knowledge.

(d) **Authentication Header** – The authentication extension header has two purposes – it validates the originator of a packet, and it ensures the integrity of the actual data carried in the packet by preventing address spoofing and connection theft. The recipient of a message must have some assurance that the implied sender is legitimate and not a phony. The recipient must also have some assurance that the data has not been altered in transit.

(e) **Encapsulated Security Payload Header** – The encapsulated (encrypted) security payload (ESP) extension header provides a means of providing confidentiality to a packet by encrypting the data, thus preventing intruders from snooping. This header must always be the last header of the header chain, as it indicates the beginning of encrypted data.

(f) **Destination Options Header** – The destination options header is used by a source to pass information to the destination without giving intermediate routers access to the information. The format for the destination options header is the same as the hop-by-hop option except the type numbers are different. The hop-by-hop and destination options separate the set of options that should be examined at each hop (router) from those that should be interpreted only by the final destination. When the destination option is placed early in the packet, intermediate routers examine it, and when it appears after the ESP header, it can be examined only at the final destination.

Q.67. Write down the advantages of IPv6 over IPv4.

Ans. IPv6 has following advantages (enhancements) over IPv4 –

(i) **Expanded Address Space** – IPv6 uses 128-bit addresses instead of the 32-bit addresses of IPv4. This is an increase of address space

factor of 2^{96} . It has been pointed out that this allows on the order of 6×10^{23} unique addresses per square meter of the surface of the earth. Even if addresses are very inefficiently allocated, this address space seems secure.

(ii) **Improved Option Mechanism** – IPv6 options are placed in separate optional headers that are located between the IPv6 header and the transport layer header. Most of these optional headers are not examined or processed by any router on the packet's path. This simplifies and speeds up computer processing of IPv6 packets compared to IPv4 datagrams. It also makes it easier to add additional options.

(iii) **Address Autoconfiguration** – This compability provides for dynamic assignment of IPv4 addresses.

(iv) **Increased Addressing Flexibility** – IPv6 includes the concept of anycast address, for which a packet is delivered to just one of a set of nodes. The scalability of multicast routing is improved by adding a scope field to multicast addresses.

(v) **Support for Resource Allocation** – Instead of the type-of-service field in IPv4, IPv6 enables the labeling of packets belong to a particular traffic flow for which the sender requests special handling. This aids in the support of specialized traffic such as real-time video.

(vi) **Support for More Security** – The encryption and authentication options in IPv6 provide confidentiality and integrity of the packet.

(vii) **Better Header Format** – IPv6 uses a new header format in which options are separated from the base header and inserted, when needed, between the base header and the upper-layer data. This simplifies and speeds up the routing process because most of the options do not need to be checked by routers.

Q.68. Write brief note on IPv6 addressing.

Ans. An IPv6 address uniquely identifies a computer on the Internet. The basic difference between IPv6 addressing and IPv4 addressing is the address length. Whereas IPv4 works with 32-bit IP addresses, IPv6 uses 128-bit IP addresses, making the address space extremely large. IPv6 address can be one of three types –

(i) **Unicast Address** – A single computer is identified by this address. For an unicast address, a datagram heading is routed to the destination computer with the shortest possible path.

(ii) **Anycast Address** – A set of computers is identified by this address, which share a common prefix. Although, the datagram is delivered only to one of the set. A packet transmit to an anycast address is delivered to one of the interfaces identified by that address.

(iii) **Multicast Address** – A set of computers is identified by address, each one of which can be at a different physical location. If a datagram known as multicast address, a copy of that datagram is delivered to each of the computers which are mapped to a multicast address.

Q.69. Explain the various features introduced by IPv6.

Ans. The major features (changes) introduced by IPv6 can be grouped into seven categories –

(i) **Larger Addresses** – The new address size is the most noticeable change. IPv6 quadruples the size of an IPv4 from 32 to 128 bits.

(ii) **Extended Address Hierarchy** – IPv6 uses the larger address space to create additional levels of addressing hierarchy.

(iii) **Flexible Header Format** – IPv6 uses an entirely new and incompatible datagram format. Unlike the IPv4 fixed-format header, IPv6 defines a set of optional headers.

(iv) **Improved Options** – Like IPv4, IPv6 allows a datagram to include optional control information. IPv6 includes new options that provide additional facilities not available in IPv4.

(v) **Provision for Protocol Extension** – Perhaps the most noticeable change in IPv6 is a move from a protocol that fully specifies all details to a protocol that permit additional features. The extension capability has the potential to allow the IETF to adopt the protocol to changes in underlying network hardware or to new applications.

(vi) **Support for Autoconfiguration and Renumbering** – IPv6 provides facilities that allow computers on an isolated network to assign themselves addresses and begin communicating without depending on a router or manual configuration. The protocol also includes a facility that permits a network manager to renumber networks dynamically.

(vii) **Support for Resource Allocation** – IPv6 has two facilities that permit preallocation of network resources, a flow abstraction and differentiated service specification.

UNIT

2

STUDY OF TRADITIONAL ROUTING AND TRANSPORT

ROUTING PROTOCOLS – BGP – CONCEPT OF HIDDEN NETWORK AND AUTONOMOUS SYSTEM, AN EXTERIOR GATEWAY PROTOCOL, DIFFERENT MESSAGES OF BGP, INTERIOR GATEWAY PROTOCOL – RIP, OSPF, MULTIPLEXING AND PORTS

Q.1. Define routing. Explain intra and inter domain routing.

Ans. Routing is one of the most important functions of IP. Fig. 2.1 shows a simplified view of the processing done at the IP layer. Datagrams to be routed can be generated either on the local host or on some other host. In the latter case this host must be configured to act as a router, or datagrams received through the network interfaces that are not ours are dropped. (i.e., silently discarded).

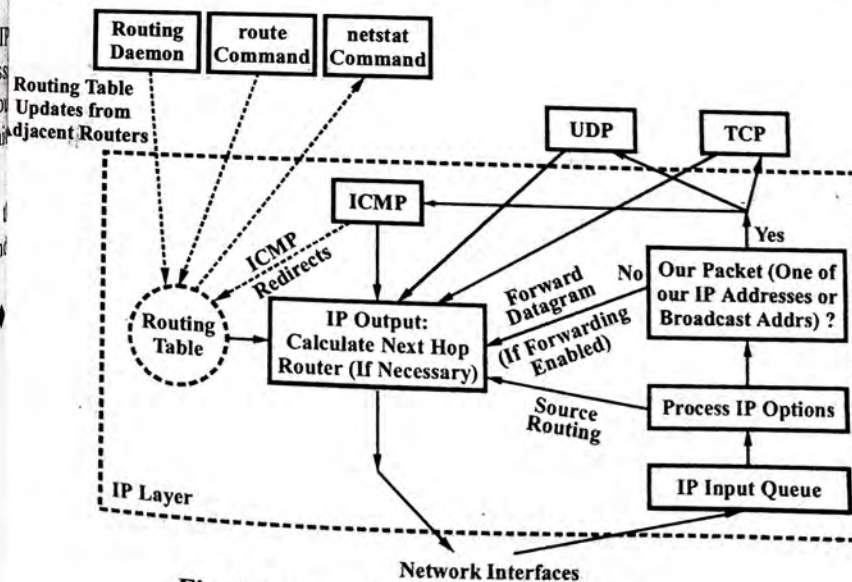


Fig. 2.1 Processing Done at the IP Layer

In fig. 2.1 we show a routing daemon, which is normally a user process. The routing table that we show in fig. 2.1 is accessed frequently by IP but updated much less frequently by a routing daemon. The routing table can also be updated when ICMP "redirect" messages are received. The route command is executed when the system is bootstrapped, to install some initial routes. It will also use the netstat command to display the routing table.

In a system such as the Internet, many different routing protocols are currently used. The Internet is organized into a collection of *autonomous systems (ASs)*, each of which is normally administered by a single entity. A corporation or university campus often defines an autonomous system. The NSFNET backbone of the Internet forms an autonomous system, because the routers in the backbone are under a single administrative control.

Each autonomous system can select its own routing protocol to communicate between the routers in that autonomous system. This is called *interior gateway protocol (IGP)* or *intradomain routing protocol*. The most popular IGP has been the routing information protocol (RIP). A newer IGP is the open shortest path first protocol (OSPF). It is intended as a replacement for RIP. An older IGP that has fallen out of use is HELLO – the IGP used in the original NSFNET backbone in 1986.

Separate routing protocols called *exterior gateway protocols (EGPs)* or *interdomain routing protocols* are used between the routers in different autonomous systems. Historically the predominant EGP has been a protocol with the same name – EGP. A newer EGP is the border gateway protocol (BGP) that is currently used between the NSFNET backbone and some of the regional networks that attach to the backbone. BGP is intended to replace EGP.

Q.2. Discuss the concept of hidden networks with the help of an example.

Ans. An example, that will show the concept of hidden network, is shown in fig. 2.2. Here, a site has multiple LANs with multiple routers connecting them. We consider that the site has just installed local network 4 and has obtained an Internet address for it. We also consider that the routers R₂, R₃, and R₄ each have correct routes for all four of site's local networks as well as a default route that passes other traffic to the ISP's router R₁. Hosts directly connected to local network 4 may communicate with one another, and any computer on that network may route packets out to other Internet sites. However, router R₁ does not know about local network 4, because it connects only to local network 1. We say that, local network 4 is hidden behind local network 1, from the viewpoint of the ISP's routing system.

Routing information should flow from a group of participating routers to a nonparticipating router, and a nonparticipating router should pass information about hidden networks to the group. Ideally, a single mechanism should solve both problems.

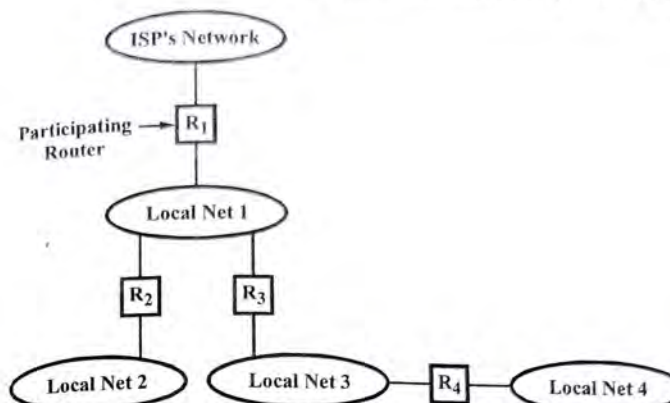


Fig. 2.2 An Example of Multiple Networks and Routers with a Single Backbone Connection

Router R₄ is the router most closely associated with local network 4, but it lies 2 hops away from the nearest core router. Hence, R₄ must depend on router R₃ to route packets to network 4. Notice that, R₄ knows about local network 4 but cannot pass datagrams directly from R₁. Router R₃ lies one hop from the core and can guarantee to pass packets, but it does not directly connect to local network 4. So, it seems incorrect to grant R₃ responsibility for advertising network 4.

Q.3. Discuss the concept of autonomous system in detail.

Ans. For purposes of routing, a group of networks and routers controlled by a single administrative authority is known as an autonomous system. Routers within an autonomous system are free to choose their own mechanisms for discovering, propagating, validating, and checking the consistency of routes. Note that, the original Internet core routers formed an autonomous system. Each change in routing protocols within the core autonomous system was made without affecting the routers in other autonomous systems.

Fig. 2.3 shows the architecture of an Internet with autonomous systems at backbone sites. To make networks that are hidden inside autonomous systems reachable throughout the Internet, each autonomous system must advertise its networks to other autonomous systems. An advertisement can be transmitted to any autonomous system. In a centralized, core architecture, however, it is crucial that each

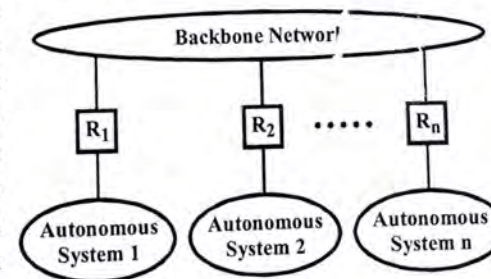


Fig. 2.3

autonomous system propagate information to one of the routers in the autonomous system.

An autonomous system owned by a corporation may choose not to route packets through an autonomous system owned by another even though they connect directly. To make it possible for automated routing algorithms to distinguish among autonomous systems, each is assigned an autonomous system number by the central authority that is charged with assigning all Internet network addresses. When routers in two autonomous systems exchange routing information, the protocol arranges for messages to carry the autonomous system number of the system each router represents.

We said that an autonomous system requires to collect information about all its networks and designate one or more routers to pass the information to other autonomous systems.

Q.4. Write short note on an exterior gateway protocol.

Ans. The term exterior gateway protocol is used to represent any protocol using pass routing information between two autonomous systems. Currently, a single exterior protocol is used in most TCP/IP Internets, known as the BGP (border gateway protocol).

When a pair of autonomous systems agree to exchange routing information, each must designate a router that will speak BGP on its behalf. The two routers are said to become BGP peers of one another. Because a router speaking BGP must communicate with a peer in another autonomous system, it makes sense to choose a machine that is near the "edge" of the autonomous system. Hence, BGP terminology calls the machine a border gateway or border router. Fig. 2.4 shows the idea.

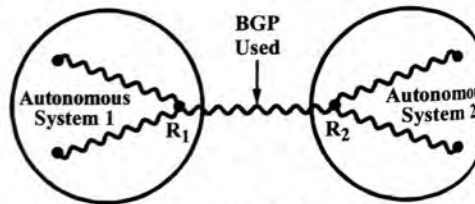


Fig. 2.4

In fig. 2.4, router R₁ gathers information about networks in autonomous system 1 and reports that information to router R₂ using BGP, while router R₂ reports information from autonomous system 2.

Q.5. What is border gateway protocol (BGP) ?

Ans. BGP is an exterior gateway protocol for communication between routers in different autonomous systems. It is a replacement for the older EGP that was used on the ARPANET. BGP version-3 is described in RFC 1266 [Lougheed and Rekhter 1991]. RFC 1268 [Rekhter and Gross 1991] describes the use of BGP in the Internet. Also, during 1993 BGP version-4 was under development, to support CIDR.

A BGP system exchanges network reachability information with other BGP systems. This information includes the full path of autonomous systems that traffic must transit to reach these networks. This information is adequate to construct a graph of AS connectivity. Routing loops can then be pruned from this graph and routing policy decisions can be enforced.

We first categorize an IP datagram as an AS either local traffic or transit traffic.

Local traffic in an AS either originates or terminates in that AS. That is, either the source IP address or the destination IP address identifies a host in that AS. Anything else is called transit traffic.

An AS can be categorized as one of the following –

- (i) A stub AS has only a single connection to one other AS. A stub AS carries only local traffic.
- (ii) A multihomed AS has connections to more than one other AS, but refuses to carry transit traffic.
- (iii) A transit AS has connections to more than one other AS and is designed under certain policy restrictions to carry both local and transit traffic.

BGP allows for policy-based routing. Policies are determined by the AS administrator and specified to BGP in configuration files. Policy decisions are not part of the protocol, but policy specifications allow a BGP implementation to choose between paths when multiple alternatives exist and to control the redistribution of information. Routing policies are related to political, security, or economic considerations.

Q.6. Describe various types of BGP messages (or packets).

Ans. BGP uses four types of messages – open, update, keepalive and notification. Fig. 2.5 shows the types of BGP messages.

The explanation of types of messages are as follows –



Fig. 2.5 Types of BGP Messages

(i) **Open Message** – Open message is used to create a neighbourhood relationship. If the neighbour accepts the neighbourhood relationship, it responds with a keepalive message, which means that a relationship has been established between the two routers. Fig. 2.6 shows the open message format.

The fields of the open message are as follows –

- (a) **Version** – The 2-byte field tells the version of BGP. The current version of BGP is 4.
- (b) **My Autonomous System** – This 2-byte field defines the autonomous system number.

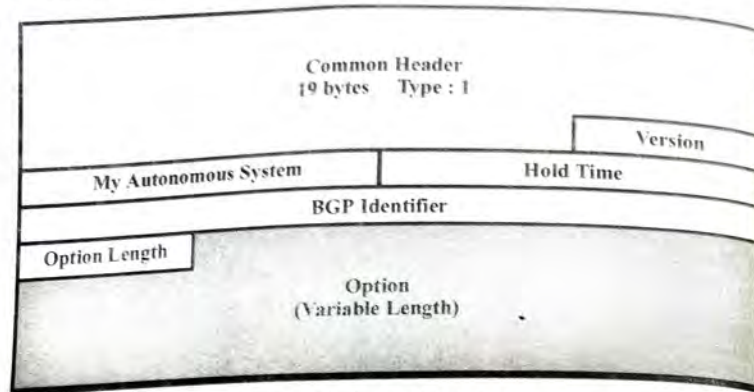


Fig. 2.6 Open Message

(c) **Hold Time** – This 2-byte field defines the maximum number of seconds that can elapse before one of the parties get a keepalive or update message from the other. If such a message is not received by the router during the holding period, it considers the other party dead.

(d) **BGP Identifier** – This is a 4-byte field defining the router that sends the open message. For this purpose, generally the router uses its ID of the IP addresses.

(e) **Option Parameter Length** – This 1-byte field defines the length of the total option parameters. This field is 0, if there are no option parameters.

(f) **Option Parameters** – Each option parameter has two subfields – the length of the parameter and the parameter value. The first option defined so far is authentication.

(ii) **Update Message** – This is the heart of the BGP protocol. It is used by a router to withdraw destinations that have been advertised previously, announce a route to new destination or both. Fig. 2.7 shows the format of an update message.

The various fields of update message are as follows –

(a) **Unfeasible Routes Length** – This 2-byte field defines the length of the next field.

(b) **Withdrawn Routes** – This field gives a list of all the routes that should be deleted from the previously advertised list.

(c) **Path Attributes Length** – This 2-byte field tells the length of the next field.

(d) **Path Attributes** – This field tells the attributes of the route to the network whose reachability is being announced in this message.

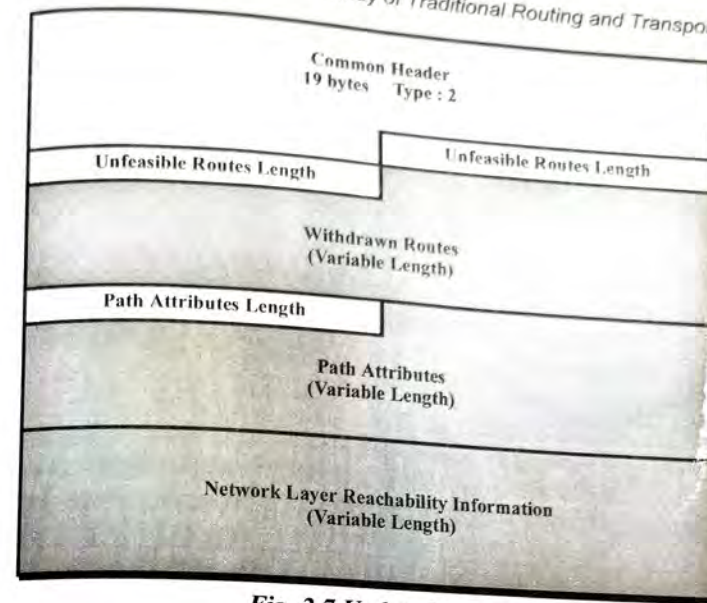


Fig. 2.7 Update Message

(e) **Network Layer Reachability Information (NLRI)** – This field defines the network that is actually advertised by this message. It has two subfields – length and IP address prefix. The length defines the number of bits in the prefix. The prefix defines the common part of the network address.

(iii) **Keepalive Message** – This message is used to tell each of the routers that they are alive. Routers exchange this information to tell each other of their availability. The keepalive message consists of only the common header as shown in fig. 2.8.



Fig. 2.8 Keepalive Message

(iv) **Notification Message** – A notification message is sent by a router whenever an error condition is detected or a router wants to close the connection. Fig. 2.9 shows the format of the notification message.

The various fields of this message are as follows –

(a) **Error Code** – This 1-byte field defines the category of the error (see table 2.1).

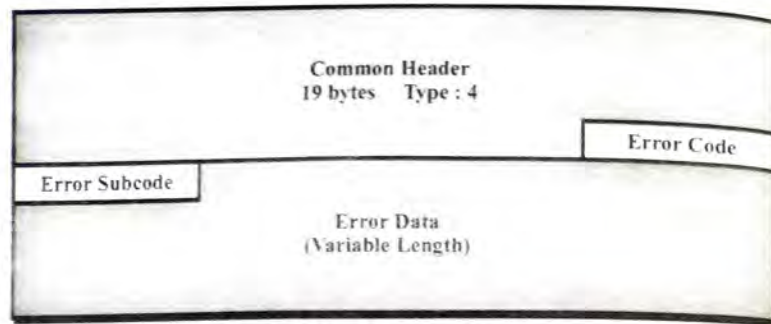


Fig. 2.9 Notification Message

Table 2.1 Error Codes

Error Code	Error Code Description	Error Subcode Description
1	Message header error	Three different subcodes are defined for this type of error—synchronization problem (1), bad message length (2), and bad message type (3).
2	Open message error	Six different subcodes are defined for this type of error—unsupported version number (1), bad peer AS (2), bad BGP identifier (3), unsupported optional parameter (4), authentication failure (5), and unacceptable hold time (6).
3	Update message error	Eleven different subcodes are defined for this type of error—malformed attribute list (1), unrecognized well-known attribute (2), missing well-known attribute (3), attribute flag error (4), attribute length error (5), invalid origin attribute (6), AS\routing loop (7), invalid next hop attribute (8), optional attribute error (9), invalid network field (10), malformed AS_PATH (11).
4	Hold timer expired	No subcode defined.
5	Finite state machine error	This defines the procedural error. No subcode defined.
6	Cease	No subcode defined.

(b) **Error Subcode** – This 1 byte field defines the type of error in each category.

(c) **Error Data** – This field can be used to give more diagnostic information about the error.

Q.7. Write the characteristics of BGP.

Ans. BGP is unusual in several ways. Most important, BGP is neither a pure distance-vector protocol nor a pure link state protocol. It can be characterized by the following –

(i) **Next Hop Paradigm** – Like distance-vector routing protocols, BGP supplies.

(ii) **Inter-autonomous System Communication** – Because BGP is designed as an exterior gateway protocol, its primary role is to allow one autonomous system to communicate with another.

(iii) **Policy Support** – Unlike most distance-vector protocols that advertise exactly the routes in the local routing table, BGP can implement policies that the local administrator chooses.

(iv) **Reliable Transport** – BGP is unusual among protocols that pass routing information because it assumes reliable transport. Thus, BGP uses TCP for all communication.

(v) **Incremental Updates** – To conserve network bandwidth, BGP does not pass full information in each update message. Instead, full information is exchanged once and then successive messages carry incremental changes called deltas.

(vi) **Coordination among Multiple BGP Speakers** – If an autonomous system has multiple routers each communicating with a peer in an outside autonomous system, BGP can be used to coordinate among routers in the set to guarantee that they all propagate consistent information.

(vii) **Path Information** – In addition to specifying destinations that can be reached and a next hop for each, BGP advertisement include path information that allows a receiver to learn a series of autonomous systems along a path to the destination.

(viii) **Route Aggregation** – BGP conserves network bandwidth by allowing a sender to aggregate route information and send a single entry to represent multiple, related destinations.

(ix) **Authentication** – BGP allows a receiver to authenticate message.

(x) **Propagation of Reachability Information** – BGP allows an autonomous system to advertise destinations that are reachable either in or through it and to learn such information from another autonomous system.

Q.8. Discuss header format of BGP.

Ans. All BGP messages have the same common header. Fig. 2.10 shows the format of this common header.

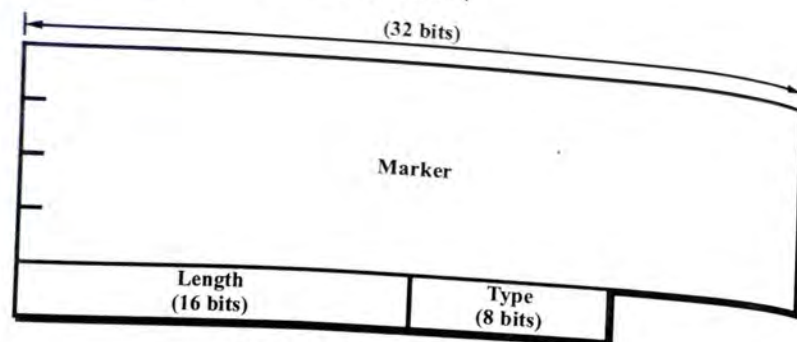


Fig. 2.10 BGP Packet Header

The fields of this header are –

- (i) **Marker** – The 16-byte marker field is reserved for authentication.
- (ii) **Length** – This 2-byte field defines the length of the total message including the header.
- (iii) **Type** – This 1-byte field defines the type of the packet. There are four types 1-4.

Q.9. Explain routing information protocol (RIP).

(R.G.P.V., Dec. 2003, June 2006, 2007)

Ans. The Routing Information Protocol (RIP) is an interior routing protocol used inside an autonomous system. RIP is developed at the University of California at Berkeley to perform routing on their local networks. Routers that connect multiple networks use RIP to let each other know the shortest route to a specified network. It is a very simple protocol based on distance vector routing which uses the Bellman-Ford algorithm for calculating the routing tables.

In distance vector routing, each router periodically shares its knowledge about the entire Internet with its neighbours. The three keys to understanding how this algorithm works, are as follows –

(i) **Sharing Knowledge About the Entire Autonomous System** – Each router shares its knowledge about the entire autonomous system with its neighbours. At the outset, a router's knowledge may be sparse.

(ii) **Sharing only with Neighbours** – Each router sends its knowledge only to neighbours. It sends whatever knowledge it has through all of its interfaces.

(iii) **Sharing at Regular Intervals** – Each router sends its knowledge to its neighbours at fixed intervals, for example every 30 seconds.

Routing Table – Every router keeps a routing table that has one entry for each destination network of which the router is aware. The entry consists of the destination network address, the shortest distance to reach the destination

in hop count, and the next hop to which the packet should be delivered to reach its final destination. The hop count is the number of the networks a packet encounters to reach its final destination. The table may be contained other information such as the subnet mask or the time this entry was last updated. Table 2.2 shows an example of a routing table.

Table 2.2 A Distance Vector Routing Table

Destination	Hop Count	Next Hop	Other Information
163.5.0.0	7	172.6.23.4	
197.5.13.0	5	176.3.6.17	
189.45.0.0	4	200.5.1.6	
115.0.0.0	6	131.4.7.19	

RIP Updating Algorithm – The routing table is updated upon receipt of a RIP response message. The following shows the updating algorithm used by RIP.

RIP Updating Algorithm

Receive : a response RIP message

1. Add one hop to the hop count for each advertised destination.
2. Repeat the following steps for each advertised destination :

If (destination not in the routing table)

Add the advertised information to the table.

Else

If (next-hop field is the same)

Replace entry in the table with the advertised one.

Else

If (advertised hop count smaller than one in the table)

Replace entry in the routing table.

3. Return.

In fig. 2.11 a router receives a RIP message from router C. The message lists destination networks and their corresponding hop counts. The first step in the RIP updating algorithm is to increase the hop count by one. Next, this updated RIP packet and the old routing table are compared. The result is a routing table with an up-to-date hop count for each destination. There is no new information for Net1, so the Net1 entry remains the same.

For Net2, information in the table and in the message identify the same next hop (i.e., router C). But the value of the hop count in the table (2) is less than the value in the message (5), the algorithm selects the one received in the message since the original value has come from router C. This value is now invalid because router C is advertising a new value.

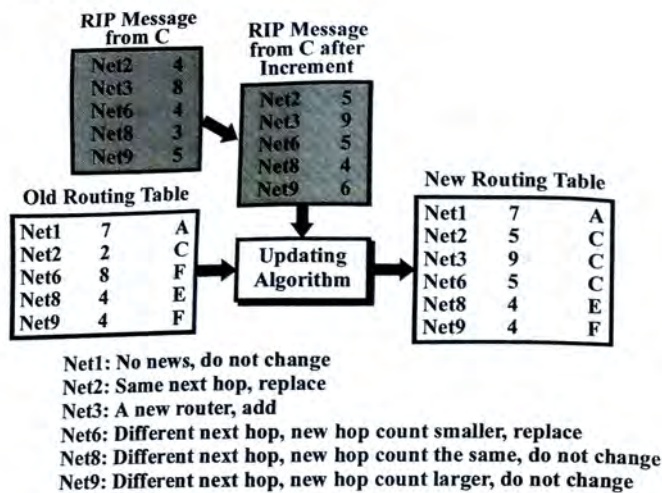


Fig. 2.11 Example of Updating a Routing Table

Net3 is added as a new destination. For Net6, the RIP packet contains lower hop count and this shows up on the new routing table. Net8 and Net9 retain their previous values because the corresponding hop counts in the message are not an improvement.

Initializing the Routing Table – Whenever a router is added to a network it initializes a routing table for itself using its configuration file. The table contains only the directly attached networks and the hop counts, which are initialized to 1. The next-hop field, which identifies the next router, is empty. Fig. 2.12 shows the initial routing tables in a small autonomous system.

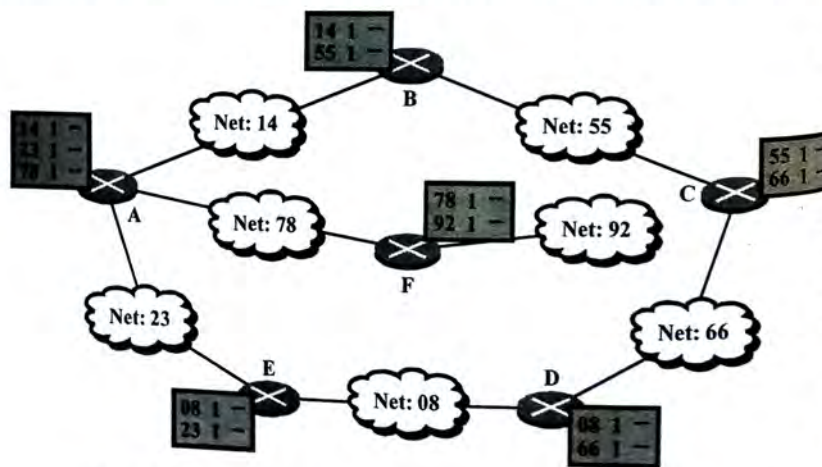


Fig. 2.12 Initial Routing Tables in an Autonomous System

Updating Routing Table – Each routing table is updated upon receipt of RIP message using RIP updating algorithm. Fig. 2.13 shows the previous autonomous system with final routing tables.

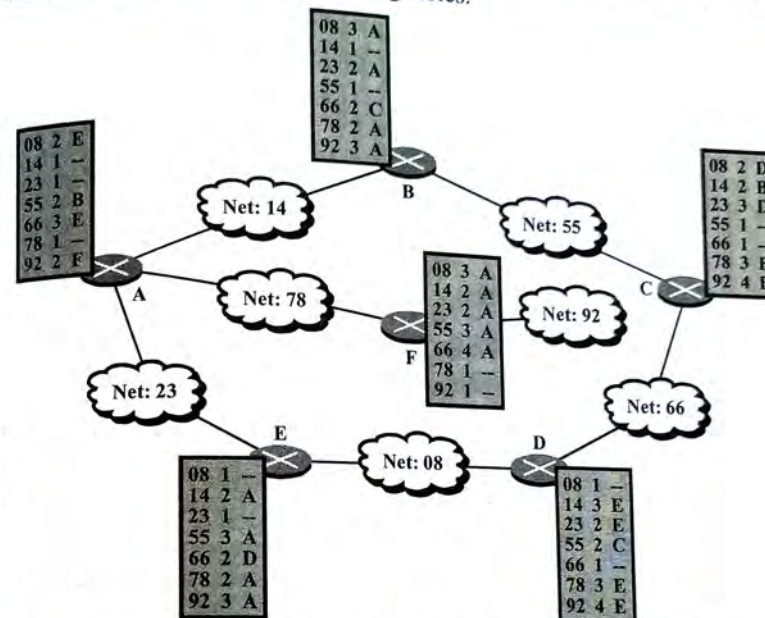


Fig. 2.13 Final Routing Table for the Previous Figure

Q.10. Discuss in brief the operation involved in RIP and the problems associated with it.

Ans. Operation – The well-known port number for RIP is UDP port 520. The normal operation of *routed*, using RIP is given below –

(i) **Initialization** – When the daemon begins it determines all the interfaces are up and sends a request packet out each interface, asking for the other router's complete routing table. On a point-to-point link this request is sent to the other end. The request is broadcast if the network supports it. The destination UDP port is 520 (the routing daemon on the other router).

This request packet has a command of 1 but the address family is set to 0 and the metric is set to 16. This is a special request that asks for a complete routing table from the other end.

(ii) **Request Received** – If the request is the special case, then the entire routing table is sent to the requestor. Otherwise each entry in the request is processed – if we have a route to the specified address, set the metric to our value, else set the metric to 16. (A metric of 16 is a special value called "infinity" and means we do not have a route to that destination). The response is returned.

(iii) **Response Received** – The response is validated and may update the routing table. New entries can be added, existing entries can be modified or existing entries can be deleted.

(iv) **Regular Routing Updates** – All or part of the router's entire routing table is sent to every neighbour router in every 30 seconds. The routing table is either broadcast (e.g., on an Ethernet) or sent to the other end of a point-to-point link.

(v) **Triggered Updates** – These occur whenever the metric for a route changes. The entire routing table not be sent only those entries that have changed must be transmitted.

Each route has a timeout associated with it. If a system running RIP finds a route that has not been updated for 3 minutes, that route's metric is set to infinity (16) and marked for deletion. It means we have missed six of the second updates from the router that advertised that route. The deletion of a route from the local routing table is delayed for another 60 seconds to ensure the invalidation is propagated.

Problems – There are various problems with RIP. First, RIP has no knowledge of subnet addressing. For example, if the normal 16-bit host ID is a class B address in nonzero RIP can't tell if the nonzero portion is a subnet ID or if the IP address is a complete host address. Some implementations use a subnet mask of the interface through which the RIP information arrived, which is not always correct.

Next, RIP takes a long time to stabilize after the failure of a router or link. The time is usually measured in minutes. During this setting time routing loops can occur. There are many subtle details in the implementation of RIP that must be followed to help prevent routing loops and to speed convergence. RFC 1058 contains many details on how RIP should be implemented.

The use of the hop count as the routing metric omits other variables that should be taken into consideration. Also, a maximum of 15 for the metric limits the sizes of networks on which RIP can be used.

Q.11. Explain the instability problem of RIP with an example.

Ans. An important problem with RIP is instability. This means that a network running RIP can become unstable. When this happens a packet can go from one router to another in a loop. Limiting the number of hops to 15 can improve stability but will not eliminate all of the problems.

To understand this problem, assume that the connection to Net1 in fig. 2.14 is nonfunctioning. Router A shows a cost of 1 for this network in its routing table. Router B, which can access Net1 only through router A, shows a cost of 2. When access to Net1 fails, router A immediately responds by changing the Net1 cost column to 16 (infinity). However, it may have to wait

up to 30 seconds before it can send its update with this new information. In the meantime, it could happen that router B sends its own update message to A. Thus, router A has two entries for Net1 – from its own table the cost is 16 and from router B the cost is 2. A is fooled into thinking that there is a backdoor access to Net1 through B. Router A then changes the cost column for Net1 to 3 (2 + 1) and this update gets sent to B. Router B has now two cost values for Net1, 3 (from A) and 2 (from itself). Router B knows that Net1 is accessible only through router A so it disregards its own lower cost and changes its cost to 4 (3 + 1). This updating continues until both routers reach a cost of 16. At this point, the routers realize there is no access to the network Net1.

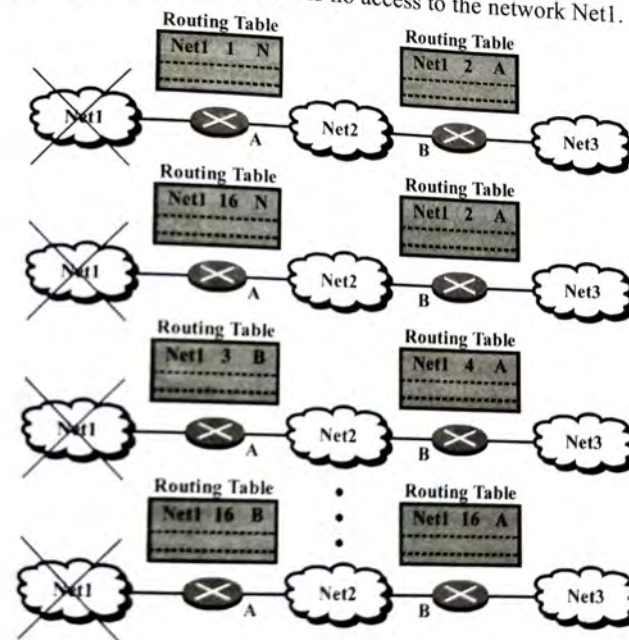


Fig. 2.14 Instability

Q.12. Discuss some remedies for instability problem with RIP.

Ans. Some remedies have been proposed to improve the stability. However, none of them are 100 percent effective.

(i) **Triggered Update** – If there are no changes on the network, updates are sent at the usual 30 seconds intervals. However, if there is a change, the router springs into action immediately by sending out its new table, a process known as **triggered update**.

Triggered update improves stability. Each router that receives an update with a change sends out new information at once, in considerably less time

than the 15 seconds average. For instance, in fig. 2.14, when router A realizes that Net1 is unavailable, it changes the cost to 16 in its routing table and immediately sends this to B. Router B then changes its table, and now both tables show a cost of 16 for Net1. The sending of update messages with incremental changes in cost has been avoided as have any looping problems.

Triggered update can improve routing, but it cannot solve all routing problems such as router failure.

(ii) **Split Horizons** – This method utilizes selectivity in the sending of routing messages to improve stability. A router must distinguish between different interfaces. If a router has received route updating information from an interface, then this same updated information must not be sent back through this interface. If an interface has passed information to help update a route, this updated information must not be sent back. It is already known and thus not required. Fig. 2.15 shows this concept. In this figure, router B has received information about Net1 and Net2 through its left interface. This information is updated and passed on through the right interface but not to the left. Similar information received by router B about Net3 is updated and passed on through the left interface of B.

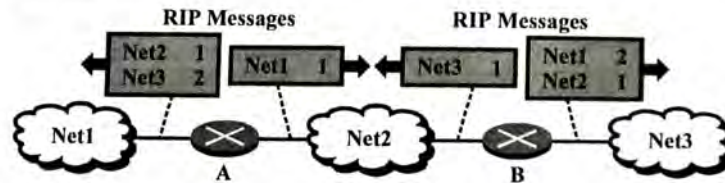


Fig. 2.15 Split Horizon

Split horizon can definitely improve stability. Suppose that Net1 is inaccessible to router A in fig. 2.14. Router B receives its information about Net1 from A. It does not send information about Net1 to router A. Therefore router A has one entry for the Net1 cost (16), and is not fooled into thinking that there is some backdoor access to Net1. Router A sends its routing table to B and both will then end up with a cost of 16 for Net1.

(iii) **Poison Reverse** – It is a variation of split horizons. In this method, information received by the router is used to update the routing table and then passed out to all interfaces. However, a table entry that has come through an interface is set to a metric of 16 as it goes out through the same interface.

Fig. 2.16 shows this concept. In this figure, router B has received information about Net1 and Net2 through its left interface, so if sent out information about these networks with a metric of 16. Similarly, information about Net3 comes from the right interface, and the cost of Net3 in the update message going right is 16.

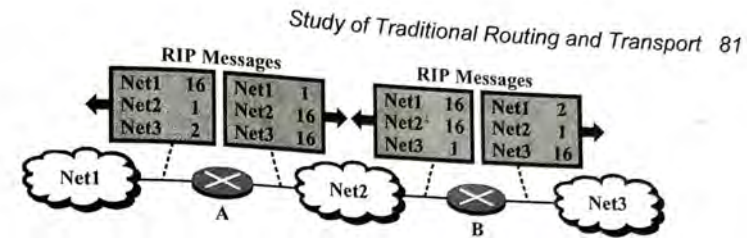


Fig. 2.16 Poison Reverse

Stability can be improved using poison reverse. Suppose that Net1 is inaccessible in fig. 2.14. Router B receives its information about Net1 from A. In each update, B sends its routing table to A with a cost of 16 for Net1. This has no effect on A if Net1 is accessible since router A will not select B's entry for Net1. However, if Net1 goes down, both cost values are 16 and instability is avoided.

Q.13. Discuss the various types of timers used in RIP.

Ans. Three timers are used to support the RIP operation (see fig. 2.17). The periodic timer controls the sending of messages, the expiration timer governs the validity of a route, and the garbage collection timer advertises the failure of a route.

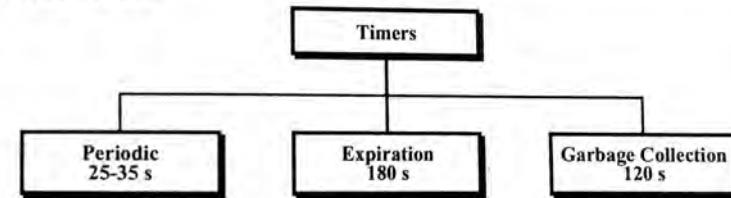


Fig. 2.17 RIP Timers

(i) **Periodic Timer** – The *periodic timer* is used to control the advertising of regular update messages. Although the protocol specifies that this timer must be set to 30 s, the working model uses a random number between 25 and 35 s. This is to prevent any possible synchronization and therefore overload on an internet if routers update simultaneously.

Each router contains one periodic timer that is set randomly to a number between 25 and 35. It counts down. When zero is reached, the update message is sent, and the timer is randomly set once again.

If RIP uses an additional timing method to send out updates, the periodic timer is not affected. The periodic update messages go out on their own schedule without regard to other update messages from other timing systems.

(ii) **Expiration Timer** – The *expiration timer* is used to govern the validity of a route. When a router receives update information for a route, the expiration timer is set to 180 s for that particular route. Every time a new

update for the route is received, the timer is reset. In normal situations this occurs every 30 s. However, if there is a problem on an Internet and no update is received within the allotted 180 s, the route is considered expired and the hop count of the route is set to 16, it means the destination is unreachable. Every route has its own expiration timer.

- (iii) **Garbage Collection Timer** – When the information about a route becomes invalid, the router does not immediately purge that route from its table. Instead, it continues to advertise the route with a metric value of 16. At the same time, a timer known as the **garbage collection timer** is set to 120 s for that route. When the count reaches zero, the route is purged from the table. This timer allows neighbours to become aware of the invalidity of a route prior to purging.

Q.14. Write a short note on open shortest path first protocol (OSPF).
(R.G.P.V., Dec. 2003, 2004, 2006)

Ans. OSPF is a newer alternative to RIP as an interior gateway protocol. This overcomes all the limitations of RIP. OSPF v2 is described in RFC1247. Its domain is also an autonomous system. Special routers called autonomous system boundary routers are responsible for dissipating information about other autonomous systems into the current system.

OSPF is a link-state protocol, as opposed to RIP, which is a distance-vector protocol. The term distance-vector means the messages sent by RIP contain a vector of distances. Each router updates its routing table based on the vector of these distances that it receives from its neighbours.

In a link-state protocol, a router does not exchange distances with its neighbours. Instead each router actively tests the status of its link to each of its neighbour, which then propagate it throughout the autonomous system. Each router takes this link-state information and builds a complete routing table. To handle routing efficiently and in a timely manner, OSPF divides an autonomous system into areas.

Areas – An area is a collection of networks, host and routers all contained within an autonomous system. An autonomous system, in turn can be divided into many different areas. All networks inside an area must be connected.

Routers inside an area flood the area with routing information. At the border of an area, special routers called area border routers summarize the information about the area and send it to other areas. Among the areas inside an autonomous system is a special area called the backbone, all of the areas inside an autonomous system must be connected to the backbone. In other words, the backbone serves as a primary area and the other areas as the secondary areas. This does not mean that the routers within areas cannot be connected with each other, however each area has an area identification. The area identification of the backbone is zero.

OSPF works by exchanging information between adjacent routers, which is not the same as between neighbouring routers. It is inefficient to have every router on a LAN talk to every other router on the LAN. Thus, one router is elected as the designated router. It is said to be adjacent to all the other routers on its and exchanges information with them. Neighbouring routers that are not adjacent do not exchange information with each other. A backup designated router is always kept up to date. Each router periodically floods LINK STATE UPDATE messages to each of its adjacent routers during normal operation. This message gives its state and gives the costs used in the topological database. The flooding messages are acknowledged to make them reliable. These messages are also sent by the routers when a line goes up or down or its cost changes. DATABASE DESCRIPTION message give the sequence numbers of all the link state entries currently held by the sender. These messages are used when a line is brought up. Either partner can request link state information from the other one by using LINK STATE REQUEST messages. The result of this algorithm is that each pair of adjacent routers checks to see who has the most recent data, and new information is spread throughout the area this way. All these messages are sent as raw IP packets. Finally, we can put all the pieces together. Using flooding, each router informs all the other routers in its area of its neighbours and costs. This information allows each router to construct the graph for its areas and compute the shortest path. The backbone area does this too.

Metric – The OSPF protocol allows the administrator to assign a cost called the metric, to each route. The metric can be based on a type of service.

OSPF protocol tackles several ambitious goals as follows –

- (i) As the name implies, the specification is available in the published literature. Making it an open standard that anyone can implement without paying license fees has encouraged many vendors to support OSPF. Consequently, it has become a popular replacement for proprietary protocols.

- (ii) OSPF includes **type of service routing**. Managers can install multiple routes to a given destination, one for each priority or type of service.

- (iii) OSPF provides **load balancing**. If a manager specifies multiple routes to a given destination at the same cost, OSPF distributes traffic over all routes equally.

- (iv) To permit growth and make the network at a site easier to manage OSPF allows a site to partition its network and routers into subsets called areas.

- (v) The OSPF protocol specifies that all exchanges between routers can be authenticated. OSPF allows a variety of authentication schemes and even allows one area to choose a different scheme than another area.

(vi) OSPF includes support for host specific, subnet-specific and classless routes as well as classful network-specific routes.

(vii) To accommodate multi-access networks like Ethernet, OSPF extends the SPF algorithm.

(viii) To permit maximum flexibility, OSPF allows managers to describe a virtual network topology that abstracts away from the details of physical connections.

(ix) OSPF allows routers to exchange routing information learned from other (external) sites.

Q.15. Give features of OSPF that make it superior to RIP.

Ans. OSPF is different from RIP in that OSPF uses IP directly. That is, it does not use UDP or TCP. There are many other features that make it superior to RIP –

(i) OSPF can calculate a separate set of routes for each IP type-of-service. This means that for any destination there can be multiple routing table entries, one for each IP type-of-service.

(ii) When several equal cost routes to a destination exist, OSPF distributes traffic equally among the routes. This is called **load balancing**.

(iii) Each interface is assigned a dimensionless cost. This can be assigned based on throughput, round-trip time, reliability or whatever. A separate cost can be assigned for each IP type of service.

(iv) OSPF supports subnets, a subnet mask is associated with each advertised route. This provides a single IP address of any class to be broken into multiple subnets of various sizes. Routes to a host are advertised with a subnet mask of all one bits. A default route is advertised as an IP address of 0.0.0.0 with a mask of all zero bits.

(v) A simple authentication scheme can be used. A clear text password can be specified, similar to the RIP-2 scheme.

(vi) OSPF uses multicasting, instead of broadcasting, to reduce the load on systems not participating in OSPF.

(vii) Point-to-point links between routers do not need an IP address at each end. These are called unnumbered networks. This can save IP addresses – a scarce resource these days.

Q.16. How BGP is different from RIP and OSPF ?

Ans. BGP is different from RIP and OSPF in that BGP uses TCP as its transport protocol. Two systems running BGP establish a TCP connection between themselves and then exchange the entire BGP routing table. Incremental updates are sent as the routing table changes.

BGP is a distance vector protocol, unlike RIP which announces hops to a destination. BGP enumerates the route to each destination (the sequence of

AS number to the destination). This removes some of the problems associated with distance-vector protocols. An AS is identified by a 16-bit number.

BGP detects the failure of either the link or the host on the other end of the TCP connection by sending a keepalive message to its neighbour on a regular basis. The recommended time between these messages is 30 seconds. This application level keepalive message is independent of the TCP keepalive option.

Q.17. Discuss OSPF header format in brief.

Ans. All OSPF packets share the same common header. Fig. 2.18 shows OSPF message header format. Each OSPF message begins with a fixed, 24-octet header. The various fields are –

0	8	16	24	31
Version (1)		Type	Message Length	
Source Router IP Address				
Area ID				
Checksum		Authentication Type		
Authentication (octets 0-3)				
Authentication (octets 4-7)				

Fig. 2.18 OSPF Message Header

(i) **Version** – This 8-bit field tells the version of OSPF protocol. It is currently version 2.

(ii) **Type** – This 8-bit field tells the type of the packet. There are 1-5 types of packets.

(iii) **Message Length** – This 16-bit field defines the length of the total message including the header.

(iv) **Source Router IP Address** – This 32-bit field tells the IP address of the router that sends the packet.

(v) **Checksum** – This field is used for error detection on the whole packet excluding the authentication type and authentication data field.

(vi) **Area Identification** – This 32-bit field tells the area within which the routing takes place.

(vii) **Authentication Type** – This 16-bit field defines the authentication method used in this area. Here, two types of authentication are defined – 0 for none and 1 for password.

(viii) **Authentication** – This 64-bit field is the actual value of the authentication data. In the future, when several authentication types are defined, this field will hold the result of the authentication calculation. Presently, if the

authentication type is 0, this field is filled with 0s. If the type is 1, this field carries an eight-character password.

Table 2.3

Type	Meaning
1	Hello (test reachability)
2	Database description
3	Link status request
4	Link status update
5	Link status acknowledgement

Q.18. Discuss in brief the various types of OSPF packets formats.

Ans. The format of the hello message packet is shown in fig. 2.19.

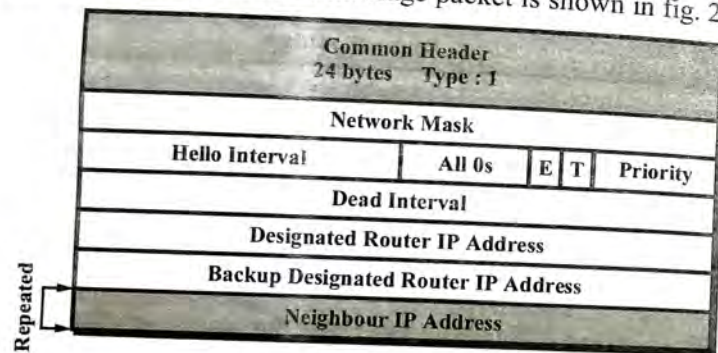


Fig. 2.19 Hello Packet

The fields are as follows –

(i) **Hello Packet** – The hello packet is used to create neighbourhood relationships and to test the reachability of neighbours.

(a) **Network Mask** – This 32-bit field defines the network mask of the network over which the hello message is sent.

(b) **Hello Interval** – This 16-bit field defines the number of seconds between hello messages.

(c) **E Flag** – It is a 1-bit field. When this flag is set, it means that the area is a stub area.

(d) **T Flag** – It is a 1-bit field. When this flag is set, it means that the router supports multiple metrics.

(e) **Priority** – It defines the priority of the router. The priority decides the selection of the designated router. Generally, the router with the highest priority is chosen as the designated router. The one with the second highest priority is chosen as the backup designated router. If the value of this field is 0, it implies that the router never wants to be a designated or backup designated router.

(f) **Designated Router IP Address** – This 32-bit field defines the IP address of the designated router for the network over which the message is sent.

(g) **Dead Interval** – This 32-bit field tells the number of seconds before a router assumes that a neighbour is dead.

(h) **Back up Designated Router IP Address** – This 32-bit field defines the IP address of the back up designated router for the network over which the message is sent.

(i) **Neighbour IP Address** – This defines the current list of all neighbours from which the sending router has received the hello message.

(ii) **Database Description Packet** – The database description packet does not contain complete database information. It only gives an outline, the title of each line in the database. The format of the database description packet is shown in fig. 2.20.

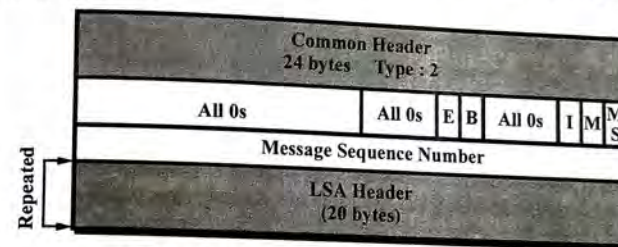


Fig. 2.20 Database Description Packet

The fields of this packet are as follows –

(a) **E Flag** – This flag is set to 1, if the advertising router is an autonomous boundary router.

(b) **B Flag** – This flag is set to 1, if the advertising router is an area border router.

(c) **M Flag** – The *more* flag is set to 1, if this is not the last message.

(d) **I Flag** – The *initialization* flag is set to 1, if the message is the first message.

(e) **M/S Flag** – The *master/slave* flag indicates the origin for the packet – master (M/S = 1) or slave (M/S = 0).

(f) **Message Sequence Number** – This 32-bit field holds the sequence number of the message. It is used to match a request with the response.

(g) **LSA Header** – It is 20-byte field used in each LSA. This header describes the outline of each link, without details.

(iii) **Link State Request Packet** – Fig. 2.21 shows the format of the link state request packet. This is a packet that is sent by a router that requires

information about a particular route or routes. It is answered with a link state update packet. This packet is used by a newly connected router to request more information about some routes after receiving the database description packet.

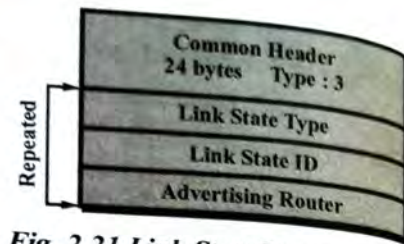


Fig. 2.21 Link State Request Packet

(iv) **Link State Update Packet** – This packet is the heart of the OSPF operation. It is used by a router to advertise the states of its links. The format of the link state update packet is shown in fig. 2.22.

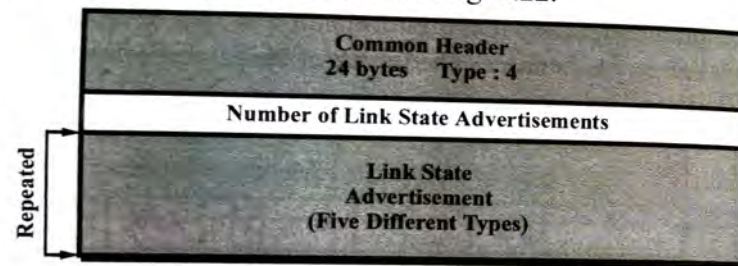


Fig. 2.22 Link State Update Packet

The fields associated with this packet are –

(a) **Number of Advertisements** – This 32-bit field defines the number of advertisements. One packet can advertise the states of several links.

(b) **Link State Advertisements** – There are five different LSAs having same header format, but different bodies.

(c) **Link State Age** – This field tells the number of seconds elapsed since this message was first generated. This message travels from router to router. When a router creates the message, the value of this field is 0. When each successive router forwards this message, it estimates the transit time and adds it to the cumulative value of this field.

(d) **E Flag** – If this flag is set to 1, it means that the area is a stub area. A stub area means a area connected to the backbone area by only one path.

(e) **T Flag** – If this flag is set to 1, it means that the router can handle multiple types of service.

(f) **Link State Type** – This field tells the LSA type. There are 5 different types.

(g) **Link State ID** – The value of this field depends on the type of link.

(h) **Advertising Router** – This is the IP address of the router advertising this message.

(i) **Link State Sequence Number** – This is a sequence number assigned to each link state update message.

(j) **Link State Checksum** – This checksum uses a special calculation called *Fletcher's checksum*, which is based on the whole packet except for the age field.

(k) **Length** – This field tells the length of the whole packet in bytes.

(v) **Link State Acknowledgement Packet** – This packet is used to acknowledge the receipt of every link state update packet to provide reliability. The format of the link state acknowledgement packet is shown in fig. 2.23. This packet has common OSPF header and the general link state update header. These two sections are enough to acknowledge a packet.

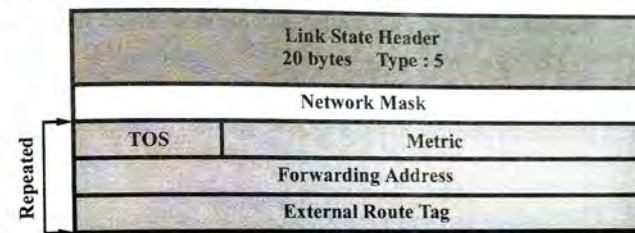


Fig. 2.23 External Link LSA

Q.19. What do you mean by multiplexing ? How it is used in data communication ? List the advantages and disadvantages of multiplexing.

Ans. In any type of network, whether it is a LAN, a MAN or a WAN the capacity or bandwidth of the transmission medium generally far exceeds that of any single attached user. Therefore, many schemes for permitting many attached users to share the larger transmission line capacity have been developed. The technique for doing this is known as **multiplexing**.

Multiplexing is the process of combining two or more signals onto a common channel. There are n inputs and a single output to a multiplexer. To make efficient use of high speed telecommunications lines, some form of multiplexing is used.

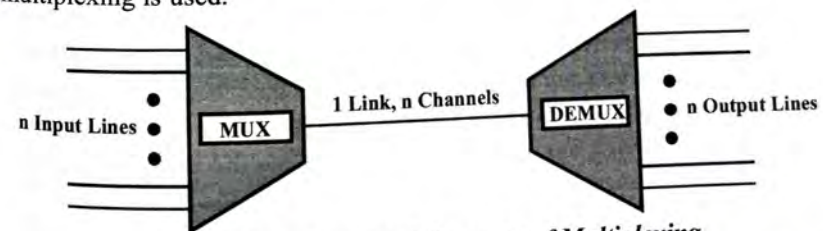


Fig. 2.24 Fundamental Concepts of Multiplexing

Multiplexing function in its simplest form is shown in fig. 2.24. There is a multiplexer having n inputs. The multiplexer is connected to a demultiplexer

by using a single data link. The link can carry n separate channels of data. Multiplexer multiplexes data from the n input lines and transmits over a higher capacity data link. Demultiplexer takes the multiplexed data stream, separates the data according to the channel and delivers them to the appropriate output lines.

In data communications, multiplexing is widely used for the following reasons –

(i) Higher the data rate, more cost-effective the transmission facility. That is, cost per kbps reduces with an increase in the data rate of the transmission facility for a given application and over a given distance. In the same way, the cost of transmission and receiving equipment, per kbps, declines with increasing the data rate.

(ii) Most individual data communicating devices need relatively modest data rate support.

Advantages of Multiplexing –

(i) It enables the use of high speed channel instead of a number of low speed channels.

(ii) It increases the bandwidth utilization.

(iii) It reduces the cable cost.

Disadvantage of Multiplexing –

It requires the extra hardware which is called multiplexer.

Q.20. What are the three major multiplexing techniques? Explain any two of them with the help of an example. (R.G.P.V., Dec. 2005)

Ans. Multiplexing can be divided in following techniques –

- Frequency division multiplexing
- Time division multiplexing
- Wavelength division multiplexing.

(i) **Frequency Division Multiplexing** – Frequency-division multiplexing is an analog technique which can be applied when the bandwidth of a link is greater than the combined bandwidths of the signals to be transmitted. In FDM, signal produced by each sending device modulates different frequencies. Then these modulated signals are combined in a single composite signal form which can be transported by the link. The bandwidth separation of carrier frequencies are kept enough to accommodate the modulated signal. These bandwidth ranges are the channels by which many signals travel. The separation of channels must be done by strips of unused bandwidth (*guard bands*) to prevent signals from overlapping. In addition, carrier frequencies must not interfere with the original data frequencies. If any of these conditions is not fulfilled, it can result in unrecoverability of the original signals.

The conceptual view of FDM is shown in fig. 2.25. In this figure, the transmission path is divided into three components, each indicating a channel to carry one transmission.

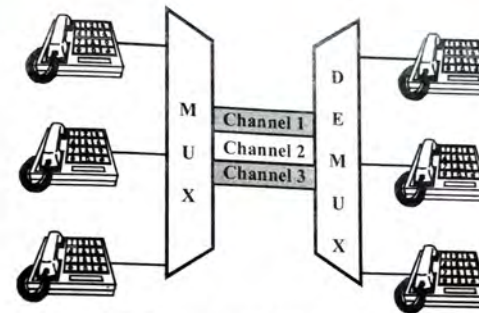


Fig. 2.25 Frequency Division Multiplexing

FDM Process – Conceptual frequency division multiplexing process in time domain is shown in fig. 2.26. FDM is an analog process and in the figure, telephones are used as the input and output devices. Each telephone produces a signal of the similar frequency range. These similar signals are modulated onto different carrier frequencies (f_1 , f_2 and f_3) inside the multiplexer. Then resulting modulated signals are combined into a single composite signal which is sent out over a media link which has sufficient bandwidth to accommodate it.

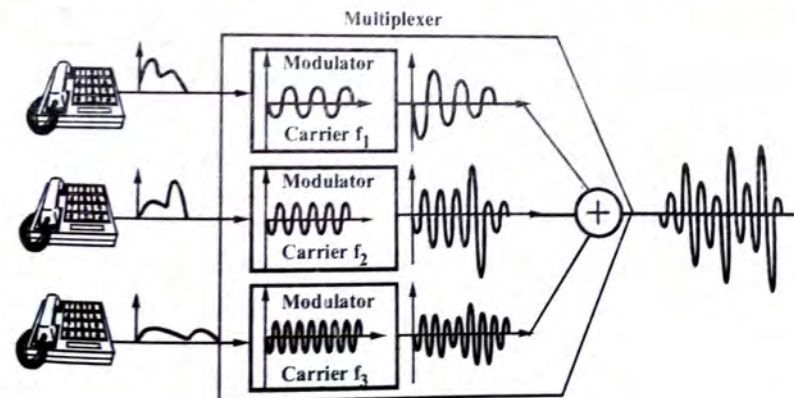


Fig. 2.26 FDM Multiplexing Process in Time Domain

(ii) **Time Division Multiplexing** – Time division multiplexing is a digital process which can be applied when the data rate capacity of the transmission medium is greater than the data rate needed by the sending and receiving devices. In this case, multiple transmissions can occupy a single link by subdividing them and interleaving the portions.

A conceptual view of TDM is shown in fig. 2.27.

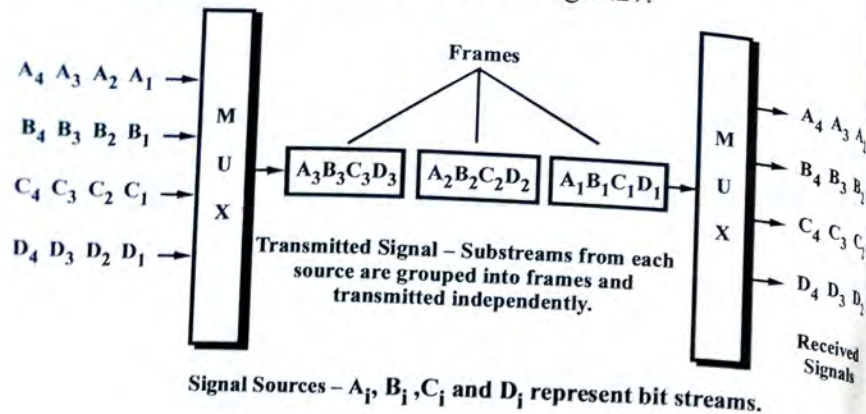


Fig. 2.27 Time Division Multiplexing

There are two types of TDM -

(a) Synchronous TDM (b) Asynchronous TDM.

(iii) **Wavelength Division Multiplexing** - Wavelength division multiplexing (WDM) is conceptually same as FDM, except that multiplexing and demultiplexing involve light signals transmitted through fiber-optic channels. Idea is same. Different signals of different frequencies are combined. But, the difference is that the frequencies are very high.

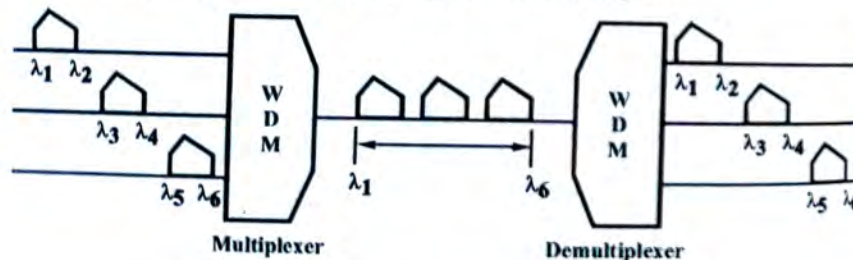


Fig. 2.28 Wavelength Division Multiplexing

A conceptual view of a WDM multiplexer and demultiplexer is shown in fig. 2.28. Very narrow bands of light from different sources are mixed to make a wider band of light. At the receiver, demultiplexer separates the signals.

Although the technology of WDM is very complex, the idea is very simple. We desire to combine multiple light sources into one single light at the multiplexer and do the reverse at the demultiplexer. A prism easily performs the combining and splitting of light sources.

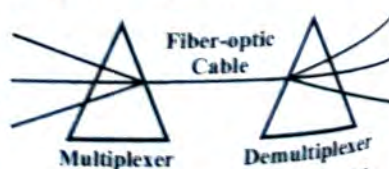


Fig. 2.29 Prisms in WDM Multiplexing and Demultiplexing

A prism bends a beam of light based on the angle of incidence and the frequency. By using this method a multiplexer can be made to combine many input beams of light, each having a narrow band of frequencies, into one output beam of a wider band of frequencies. A demultiplexer can be made to reverse the process. Concept is shown in fig. 2.29.

Q.21. Differentiate between FDM and TDM. (R.G.P.V., Dec. 2004, 2008)

Ans. The differences between FDM and TDM are as follows -

S.No.	Frequency Division Multiplexing	Time Division Multiplexing
(i)	In FDM, frequency spectrum is divided among the logical channels with each user having exclusive possession of frequency band.	In TDM, the users take turns (in a round-robin), each one periodically getting the entire bandwidth for a little burst of time.
(ii)	Frequency division multiplexing can be used with analog signals.	Time division multiplexing can be used with digital signals.
(iii)	There is no framing.	There is a basic requirement of framing as data is sent through the frames.
(iv)	FDM requires guardbands for channel separation to avoid interference.	There is no interference between the channels.
(v)	The capacity of FDM channels is divided into group, super group, master group and master super group with the band width in kilo Hz.	The capacity of TDM channels is divided into T_1, T_2, T_3, T_4 and super T_4 in Mbps.
(vi)	Examples of FDM are radio and TV systems.	Example of TDM is SONET system, ISDN channels.

Q.22. Write short note on port numbers.

Ans. A remote computer may be running several server programs at the same time. A local computer may also be running several client programs at the same time. For communication, we must define local host, local client program, remote host, remote client program. The local host and the remote host are defined using IP addresses. To define the client and server programs we need port numbers or port addresses. In TCP/IP protocol suite the port numbers are integers between 0 and 65,535. A client program running on the local computer defines itself with a port number, chosen randomly by the TCP software running on the local host. This is known as the **ephemeral port number**. The server program, running on the remote computer must also define

itself with a port number. This port number, however, cannot be chosen randomly. If the computer at the server site runs a server process and assigns a random number as the port number, the process at the client site that wants to access that server and use its services will not know the port number. TCP/IP has decided to use universal port numbers for servers, these are called **well-known port numbers**.

IANA has divided the port numbers into three ranges – well-known, registered and dynamic.

(i) **Well-known Ports** – The ports ranging from 0 to 1,023 are assigned and controlled.

(ii) **Registered Ports** – The ports ranging from 1,024 to 49,151 are not assigned and controlled by IANA. They can only be registered with IANA to prevent duplication.

(iii) **Dynamic Ports** – The ports ranging from 49,152 to 65,535 are neither controlled nor registered. They can be used by any process. These are the ephemeral (temporary) numbers.

TCP – SEGMENT FORMAT, SOCKETS, SYNCHRONIZATION, THREE WAY HAND SHAKING, VARIABLE WINDOW SIZE AND FLOW CONTROL, TIMEOUT AND RETRANSMISSION ALGORITHMS, CONNECTION CONTROL, SILLY WINDOW SYNDROME

Q.23. Write short note on TCP.

Ans. Transmission control protocol (TCP) provides a connection-oriented user-to-user byte stream service. This means it provides a logical connection between two sites and is capable of transmitting a sequence of bytes between them. It divides a byte stream into a sequence of segments and sends them to the destination via a variation on a sliding window flow control protocol. It provides the initial handshaking by establishing, maintaining, and releasing connections. It handles requests to deliver information to a destination reliably. It is an important consideration since the lower layer does not guarantee about delivery of packets. TCP receives data or requests from its user, stores it in a TCP segment format and gives it to the IP, which is shown in fig. 2.30. It plays no role in the subsequent routing and transfer of information.

Each byte on a TCP connection has its own 32-bit sequence number. In case of a host blasting away at full speed on a 10 Mbps LAN, theoretically the sequence numbers could wrap around in an hour, but practically it takes much longer. Sequence numbers are used both for acknowledgement and for the window mechanism, in which separate 32-bit header fields are used.

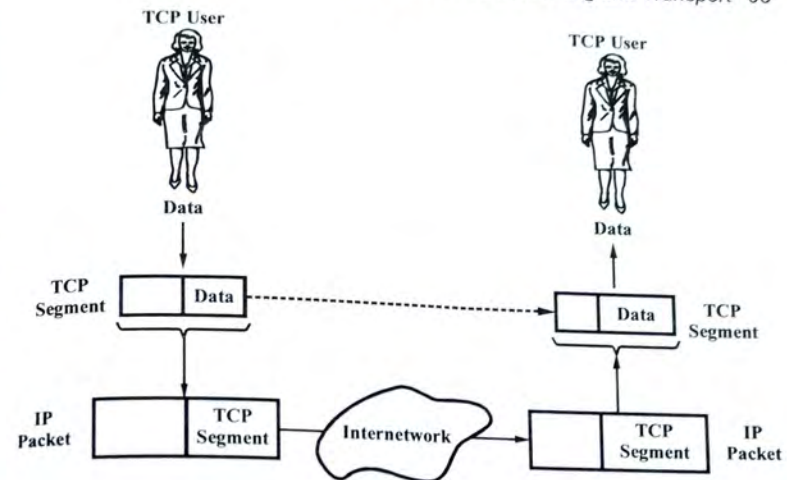


Fig. 2.30 TCP as a User-to-user Service

Q.24. Explain about the various fields of TCP header with the help of neat diagram.

Ans. The format of the TCP header is shown in fig. 2.31. Its size is varying between 20 to 60 bytes.

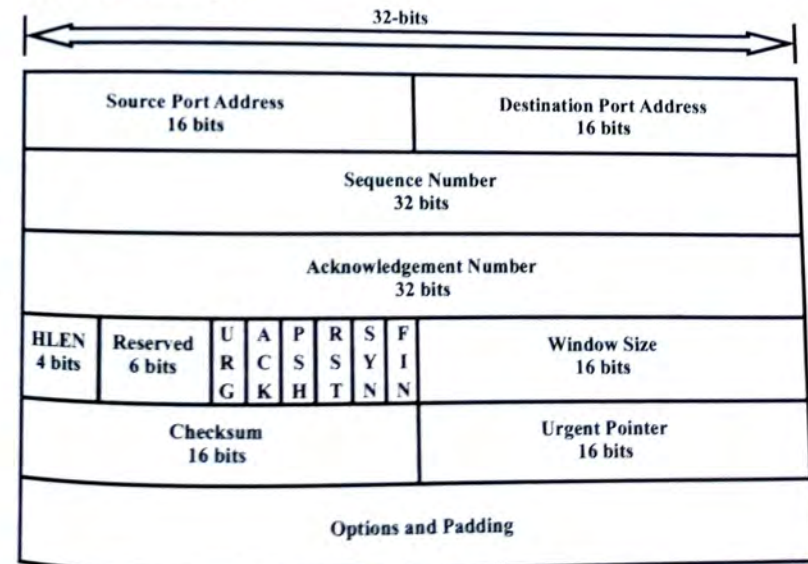


Fig. 2.31 The TCP Header

(i) **Source Port Address** – This 16-bit field specifies the port number of the application program in the host that is sending the segment.

(ii) **Destination Port Address** – This 16-bit field specifies the port number of the application program in the host that is receiving the segment.

(iii) **Sequence number** – This is a 32-bit field that specifies the number assigned to the first byte of data contained in this segment.

(iv) **Acknowledgement Number** – This is a 32-bit field that specifies the byte number that the receiver of the segment is expecting to receive from the other party.

(v) **Header Length (HLEN)** – This is a 4-bit field that tells how many 4-byte words are contained in the TCP header. The header length can be between 20 and 60 bytes. Thus the value of this field will be 5 ($5 \times 4 = 20$) and 15 ($15 \times 4 = 60$).

(vi) **Reserved** – This 6-bit field is reserved for future use.

(vii) **URG** – The URG bit is set to 1 to indicate that the value of the urgent pointer field is valid.

(viii) **ACK** – ACK is set to 1 if the value of the acknowledgement field is valid.

(ix) **PSH** – This 1-bit flag specifies PUSHed data.

(x) **RST** – This 1-bit flag is used to reset a connection that has become confused due to a host crash or some other reason.

(xi) **SYN** – This 1-bit long flag is used to establish a connection.

(xii) **FIN** – This 1-bit flag is used to release a connection. It tells that the sender has no more data to transmit.

(xiii) **Window Size** – This is a 16-bit field that specifies the size of the window, in bytes, that the other party must maintain. The length of this field is 16-bits. It means that the maximum size of the window is 65,535 bytes.

(xiv) **Checksum** – This 16-bit long field is provided for extra reliability. It checksums the header, and the data, when performing this computation, the TCP checksum field is set to zero.

(xv) **Urgent Pointer** – This 16-bit field is valid only if the urgent flag is set. This pointer is used when the segment contains urgent data.

(xvi) **Options** – This field provides a way to add extra facilities not covered by the regular header. There can be upto 40 bytes of optional information in the TCP header.

Q.25. Explain the main features of transmission control protocol.

Ans. There are following features of TCP as given below –

(i) **Numbering System** – The TCP software track of the segments being sent or received, in the segment header, there is no field to a segment number value. Instead, there are two fields known the sequence number and the acknowledgement number. These fields refer to the byte number and not a segment number.

(ii) **Byte Number** – TCP numbers all data bytes which are sent in a connection. Numbering is not dependent in every direction. It stores them in the transmitting buffer and numbers them, if TCP receives data bytes from a process. The numbering does not essentially begin from 0. TCP produces a random number between 0 to $2^{32} - 1$ to the number of the first byte.

(iii) **Sequence Number** – TCP assigns a sequence number to every segment which is being transmitted after the bytes have been numbered. For each segment, the sequence number is the number of the first byte carried in that segment.

(iv) **Acknowledgement Number** – The acknowledgement number defines the number of the next byte which the party expects to get. Furthermore, the acknowledgement number is cumulative, it means that the party takes the number of the last byte which it has received safe and sound, add 1 to it, and announces this sum as the acknowledgement number.

(v) **Error Control** – TCP implements an error control mechanism to provide reliable service. Thus error control assumes a segment as the unit of data for error detection, error control is byte-oriented.

(vi) **Flow Control** – Unlike UDP, flow control is provided by TCP. The receiver of data controls the data amount which are to be transmitted via the transmitter. This can be performed to prevent the receiver from being overwhelmed with data. To use a byte oriented flow control, the numbering system permits TCP.

(vii) **Congestion Control** – TCP takes into account congestion in the network. The data amount transmitted via a transmitter is not only handled through the receiver, but is also computed by the congestion level in the network.

Q.26. Write short note on socket addresses.

Ans. TCP or UDP needs two identifiers, the IP address and the port number, for each end to make a connection. The combination of an IP address and a port number is known as a socket address. To use the services of TCP or UDP, we need a pair of socket addresses – client socket address and server

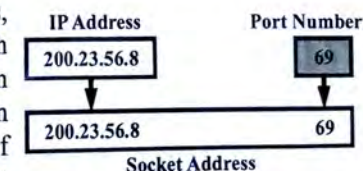


Fig. 2.32 Socket Address

socket address. The client socket address defines the client application process uniquely. The server socket address defines the server application process uniquely. These four pieces of information are part of the IP header and the TCP or UDP header. The IP header contains the IP addresses and the TCP or UDP header contains the port numbers.

Q.27. With neat diagram explain the process of connection establishment and release using 3-way handshaking in transport layer.

Ans. TCP Connection Establishment – Connections are established using the three-way handshake. To establish a connection, one side, the server, passively waits for an incoming connection by executing the LISTEN and ACCEPT primitives, either specifying a specific source or no particular.

The other side, say, the client, executes a CONNECT primitive, specifying the IP address and port to which it wants to connect, the maximum TCP segment size it is willing to accept and optionally some user data (e.g., a password). The CONNECT primitive sends a TCP segment with the SYN bit on and the ACK bit off and waits for a response.

When this segment arrives at the destination, the TCP entity there checks to see if there is a process that has done a LISTEN on the port given in the destination port field. Otherwise it sends a reply with the RST bit on to reject the connection.

If some process is listening to the port, that process is given the incoming TCP segment. Then, it can either accept or reject the connection. If it accepts, an acknowledgement segment is sent back. Fig. 2.33 (a) shows the sequence of TCP segments sent in the normal case. Note that a SYN segment consumes 1 byte of sequence space so that it can be acknowledged unambiguously.

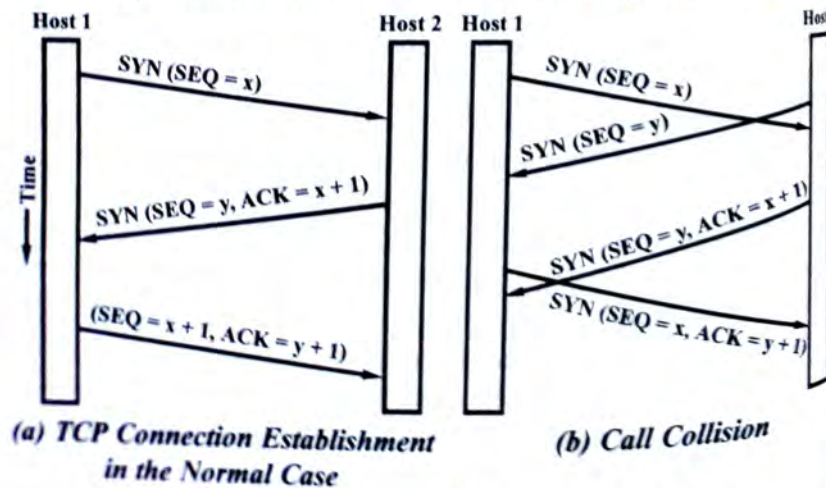


Fig. 2.33

In the event that two hosts simultaneously attempt to establish a connection between the same two sockets. This sequence of events is as shown in fig. 2.33 (b). The result of these events is that just one connection is established, not two because connections are identified by their end points. If the first setup results in a connection identified by (x, y) and the second one does too, only one table entry is made, namely, for (x, y).

A clock-based scheme is used, with a clock tick every 4 μ sec. For additional safety, when a host crashes, it may not reboot for the maximum packet lifetime (120 sec) to ensure that no packets from previous connections are still roaming around the Internet somewhere.

TCP Connection Release – However, TCP connections are full duplex, to understand how connections are released it is best to think of them as a pair of simplex connections. Each simplex connection is released independently of its sibling. To release a connection, either user can send a TCP segment with the FIN bit set, which means that it has no more data to transmit. When the FIN is acknowledged, that direction is shut down for new data. However, data may continue to flow indefinitely in the other direction. When both directions have been shut down, the connection is released. Generally, four TCP segments are needed to release a connection, one FIN and one ACK for each direction. However, it is possible for the first ACK and the second FIN to be contained in the same segment, reducing the total count to three.

Similar as with telephone calls in which both people say goodbye and hang up the phone simultaneously, both ends of a TCP connection may send FIN segments at the same time. These are each acknowledged in the usual way and the connection is shut down. In fact, there is no difference between the two hosts releasing sequentially or simultaneously.

Q.28. What is the advantage of using variable window size in TCP?

Ans. The advantage of using a variable size window is that it gives flow control as well as reliable transfer. To avoid receiving more data than it can store, the receiver sends smaller window advertisements as its buffer fills. The receiver advertises a window size of zero to stop all transmissions for extreme case. Later, when buffer space becomes available, the receiver advertises a nonzero window size to trigger the flow of data again.

Q.29. Explain flow control in TCP.

Ans. Flow control specifies the amount of data a source can send before receiving an acknowledgement from the destination. In a rare case, a transport layer protocol could send 1 byte of data and wait for an acknowledgement before sending the next byte. But this process is very slow. If the data is traveling a long distance, the source is idle while it waits for an acknowledgement.

At the other extreme case, a transport layer can send all of the data without worrying about acknowledgement. This speeds up the process, but may overwhelm the receiver. Moreover, if some part of the data is duplicated, received out of order, or corrupted, the source will not know all has been checked by the destination.

TCP provides a solution for it. It defines a window that is imposed on a buffer of data delivered from the application program and is ready to be sent. TCP sends as much data as is defined by the sliding window protocol.

Sliding Window Protocol – TCP uses a sliding window protocol to accomplish flow control. In this method, both hosts use a window for a connection. The window spans a portion of the buffer containing bytes that the host can send before worrying about an acknowledgement from the other host. The window is known as the *sliding window* as it can slide over the buffer and data and acknowledgements are sent and received.

Fig. 2.34 shows the sender buffer. In fig. 2.34, the bytes before 200 have been sent and acknowledged. Now, the sender can reuse these locations. Bytes 200 to 202 have been sent, but not acknowledged. The sender has to keep these bytes in the buffer because they can be lost or damaged. Bytes 203 to 211 are in the buffer but have not yet been sent.

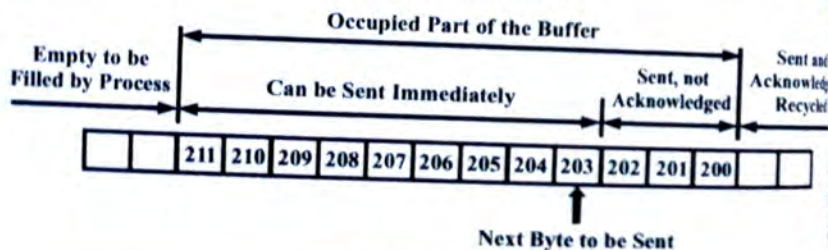


Fig. 2.34 Sender Buffer

Now assume a situation in which there is no sliding window protocol. In this case, the sender can go ahead and send all the bytes (up to 211) in the buffer, without regard to the condition of the receiver. The receiver's buffer with its limited-size could completely fill up as the receiving process is not consuming data fast enough. The excess bytes thrown away by the receiver will need retransmission. The sender must adjust itself to the number of locations present at the receiver site.

Receiver Window – The receiver buffer is shown in fig. 2.35. Now, the next byte to be consumed by the process is byte 194. The receiver expects to receive byte 200 from the sender. Generally if the total size of the receiver buffer is P and Q locations are already occupied, then only P-Q more bytes can be received. This value is known as the *receiver window*. For example, if P = 13 and Q = 6, this specifies that the value of the receiver window is 7.

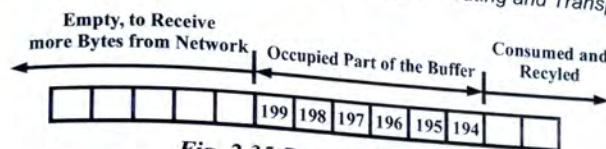


Fig. 2.35 Receiver Window

Sender Window – Flow control can be made if the sender creates a window with a size lower than or equal to the size of the receiver window. This window includes the bytes sent and not acknowledged and those that can be sent. The sender buffer with the sender window is shown in fig. 2.36.

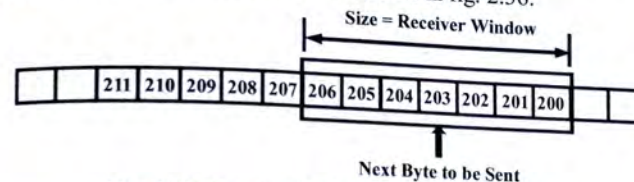


Fig. 2.36 Sender Buffer and Sender Window

Here, notice that the size of the sender window (i.e., 7) is equal to the size of the receiver window. But this does not imply that the sender can send 7 more bytes; it can send only 4 more bytes as it has already sent 3 bytes. Also note that although bytes 207 to 211 are in the sending buffer, they also cannot be sent until more news arrives from the receiver.

Sliding the Sender Window – Here, we see how messages from the receiver change the position of the sender window. Consider, in our example, that the sender sends 2 more bytes and an acknowledgement is received from the receiver (expecting byte 204) with no change in the size of the receiver window (still 7). The sender can now slide its window and the locations

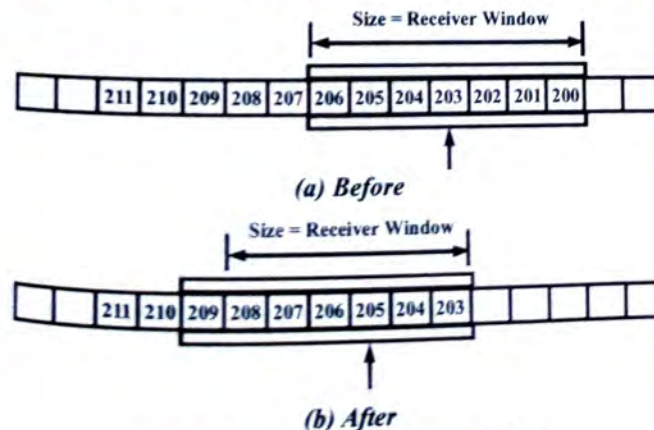


Fig. 2.37 Sliding the Sender Window

occupied by bytes 200-202 can be recycled. The position of the sender window and the sender window before and after this process (event) is shown in fig. 2.37. In the second part of the figure, the sender can now send bytes 209 (5 more bytes).

Expanding the Sender Window – The size of the receiver window expands if the receiving process consumes data faster than it receives. In fig. 2.38, the receiver has acknowledged the receipt of 2 more bytes and at the same time has increased the value of receiver window to 10. In the meantime, the sending process has created more bytes and the sending TCP has sent 5 bytes.

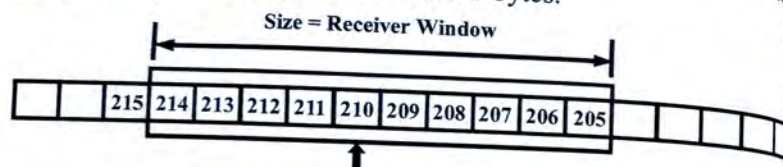


Fig. 2.38 Expanding the Sender Window

Shrinking the Sender Window – The size of the receiver window decreases if the receiving process consumes data slower than it receives. This is the case when the receiver has to inform the sender to shrink its sender window size. In fig. 2.39, the receiver has received the 5 bytes (205 to 209); however, the receiving process has consumed only 1 byte, which implies the number of free locations is reduced to 6 ($10 - 5 + 1$). It acknowledges bytes 205 to 209 (expecting 210) and also informs the sender to decrease its window size and not to send more than 6 more bytes. If the sender has already sent 2 more bytes when it receives the acknowledgment and has received 3 more bytes from the sending process, we have the following window and the buffer as shown in fig. 2.39.

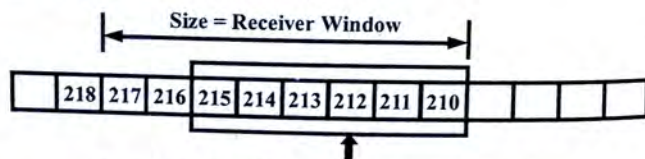


Fig. 2.39 Shrinking the Sender Window

Closing the Sender Window – If the receiver buffer is totally full, then this case the receiver window value is zero. When this is relayed to the sender, the sender closes its window (left and right walls overlap). The sender cannot send any bytes until the receiver announces a nonzero receiver window value.

Q.30. Discuss about the timeout of connection establishment.

Ans. There are several instances when the connection cannot be established. For example, the server host is down. To simulate this scenario, we issue telnet command after disconnecting the Ethernet cable from the server's host. Fig. 2.40 shows the tcpdump output.

1	0.0	bsdi.1024 > svr4, discard : S	291008001 : 291008001 (0) win 4096 <mss 1024> [tos 0 x 10]
2	5.814797 (5.8148)	bsdi.1024 > svr4, discard : S	291008001 : 291008001 (0) win 4096 <mss 1024> [tos 0 x 10]
3	29.815436 (24.0006)	bsdi.1024 > svr4, discard : S	291008001 : 291008001 (0) win 4096 <mss 1024> [tos 0 x 10]

Fig. 2.40 tcpdump Output for Connection Establishment that Times-out

The main thing in this output is how frequently the client's TCP sends a SYN to try to establish the connection. The second segment is sent 5.8 seconds after the first and the third is sent 24 seconds after the second.

Fig. 2.40 doesn't show how long the client's TCP keeps retransmitting before giving up. For seeing this, we have to time the telnet command –

bsdi % date; telnet svr4 discard; date

Thu Sep 24 16:24:11 MST 1992

Trying 140.252.13.34...

telnet : Unable to connect to remote host : Connection time out

Thu Sep 24 16:25:27: MST 1992

The time difference is 76 seconds. Most Berkeley-derived systems set a time limit of 75 seconds on the establishment of a new connection.

Q.31. How does TCP handle time out and retransmission? Explain the need for multiplexing at the transport layer.

Ans. Timeout and Retransmission – TCP wants the destination to send acknowledgements whenever it successfully receives new octets from the data stream. When it sends a segment, TCP starts a timer and waits for an acknowledgement. If the timer ends before data in the segment has been acknowledged, TCP assumes that the segment was lost or corrupted and retransmits it.

TCP contains various internet delays by using an adaptive retransmission algorithm. In other words, TCP monitors the performance of each connection and deduces reasonable values for timeouts. TCP revises its timeout value because the performance of a connection changes.

Need for Multiplexing at the Transport Layer – There is only one TCP but there are various application programs that may want to use its services in the TCP/IP protocol suite. TCP does multiplexing and demultiplexing to handle this situation.

Q.32. How does TCP manage connections in the network? Discuss thoroughly.

Ans. The steps required to establish and release connections can be represented in a finite state machine with the 11 states given in table 2.4. In

each state, certain events are legal. When a legal event happens, some action may be taken. If some other event happens, an error is occurred.

Table 2.4 The States used in the TCP Connection Management Finite State Machine

State	Description
CLOSED	No connection is active or pending
LISTEN	The server is waiting for an incoming call
SYN RCVD	A connection request has arrived; wait for ACK
SYN SENT	The application has started to open a connection
ESTABLISHED	The normal data transfer state
FIN WAIT 1	The application has said it is finished
FIN WAIT 2	The other side has agreed to release
TIMED WAIT	Wait for all packets to die off
CLOSING	Both sides have tried to close simultaneously
CLOSE WAIT	The other side has initiated a release
LAST ACK	Wait for all packets to die off

Each connection starts in the *CLOSED* state. It leaves that state when either a passive open (*LISTEN*) or an active open (*CONNECT*). If the other side does the opposite one, a connection is established and state becomes *ESTABLISHED*. Connection release can be initiated by either side. When complete the state returns to *CLOSED*.

The finite state machine is shown in fig. 2.41. The common case of a client actively connecting to a passive server is shown with heavy lines, solid for the client, dotted for the server. The lightface lines are unusual event sequences. Each line in fig. 2.41 is marked by an *event/action* pair. The event can either be a user-initiated system call (*CONNECT*, *LISTEN*, *SEND*, or *CLOSE*), a segment arrival (*SYN*, *FIN*, *ACK*, or *RST*), or, in one case, a timeout of twice the maximum packet lifetime. The action is the sending of a control segment (*SYN*, *FIN*, *RST*) or nothing, indicated by comments are shown in parentheses.

One can best understand the diagram by first following the path of an active client then later following the path of a server. When an application program on the client machine issues a *CONNECT* request, the local TCP entity creates a connection record, marks it as being in the *SYN SENT* state and sends a control segment. Notice that many connections may be open (or being opened) at the same time on behalf of multiple applications, so the state is per connection and recorded in the connection record. When the *SYN + ACK* arrives at the client, it sends the final *ACK* of the three-way handshake and switches into the *ESTABLISHED* state. Now, data can be sent and received.

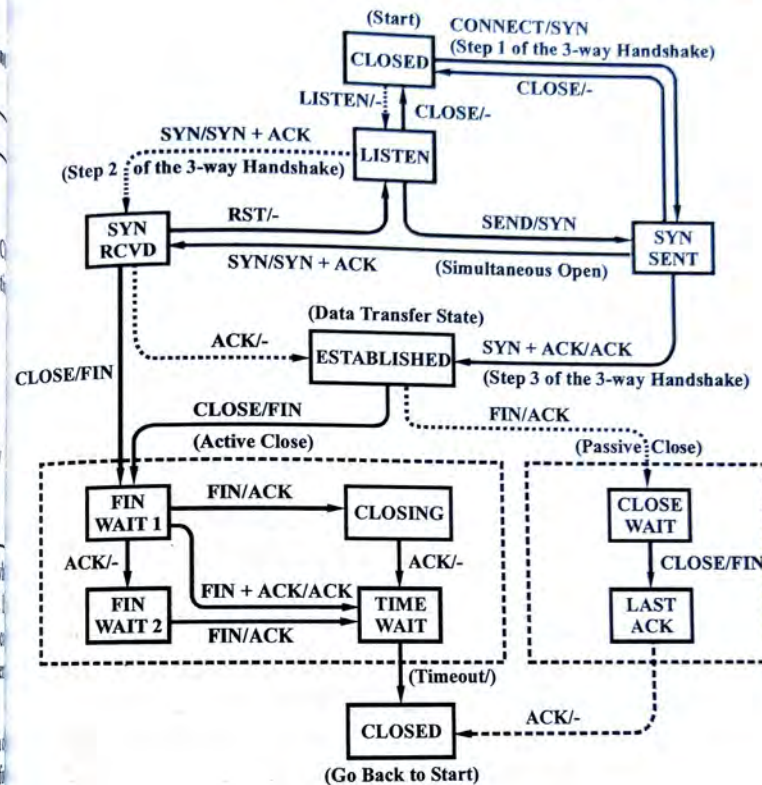


Fig. 2.41 TCP Connection Management Finite State Machine

When an application is finished, it executes a *CLOSE* primitive, which causes the local TCP entity to send a *FIN* segment and wait for the corresponding *ACK* (dashed box marked active close). When the *ACK* arrives, a transition is made to state *FIN WAIT 2* and one direction of connection is now closed. When the other side closes, too, a *FIN* comes in, which is acknowledged. Now both sides are closed, but TCP waits a time equal to the maximum packet lifetime to guarantee that all packets from the connection have died off. Similarly, in case of acknowledgement was lost. When the timer goes off, TCP deletes the connection record.

Now, let us examine connection management from the server's point of view. The server does a *LISTEN* and settles down to see who turns up. When a *SYN* comes, it is acknowledged and the server goes to the *SYN RCVD* state. When the server's *SYN* is itself acknowledged, the three-way handshake is complete and the server goes to the *ESTABLISHED* state.

Q.33. Define a term silly window syndrome and possible solution to overcome its effect.

Ans. Silly window syndrome is a problem which can degrade performance. If large block of data is passed from transmitter but the receiver reads data one byte at a time. Receiving side, the TCP buffer is full and the transmitter know the condition. The interactive application reads one character from the TCP stream. Receiving TCP tells to the transmitter to send the next byte. Transmitter send 1 byte. Now buffer is full and receive acknowledgement the 1-byte segment and set the window 0. This opens the window continuous. Fig. 2.42 shows these steps.

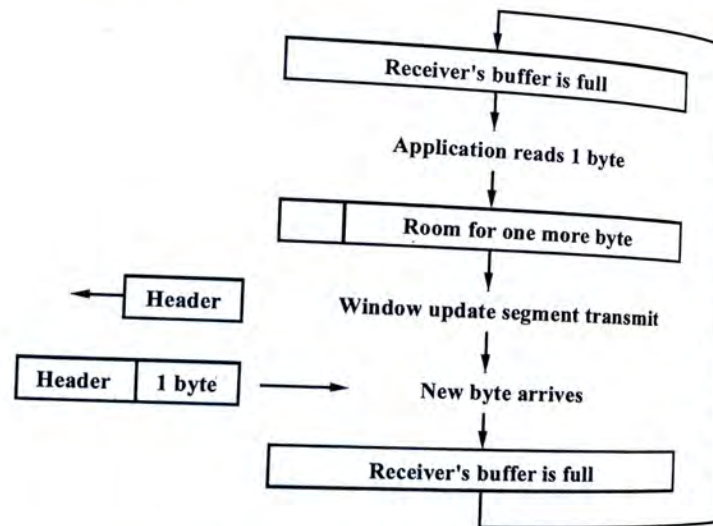


Fig. 2.42 Silly Window Syndrome

For the silly window syndrome, Nagle algorithm and Clark's solution are complementary. Clark's solution is to prevent the receiver from sending a window update for 1 byte. Instead it is forced to wait until it has a decent amount of data in its buffer space available.

Specifically, the receiver must not transmit a window update until it has a decent amount of data in its buffer space available. To manage the maximum size of segment, it advertises if the connection is established or until its buffer is half empty, whichever is smaller.

Clark was trying to solve the problem of the receiving application sending data up from TCP a byte at a time and Nagle was trying to solve the problem of the transmitting application delivering data to TCP a byte at a time. Both solutions are valid and can work together. The aim is for the transmitter not to transmit small segments and the receiver not to ask for them.

EXAMPLE OF TCP – TAHOE, RENO, SACK, ETC.

Q.34. Explain Tahoe with problem and congestion avoidance.

Ans. Tahoe refers to the TCP congestion control algorithm which was suggested by Van Jacobson. TCP packet transmissions are clocked by the receiving acknowledgements. However there is a problem when a connection first starts up cause to have acknowledgements you need to have data in the network and to put data in the network you need acknowledgements. To get around this circularity Tahoe suggests that whenever a TCP connection starts or re-starts after a packet loss it should go through a procedure called 'slow-start'. The reason for this procedure is that an initial burst might overwhelm the network and the connection might never get started. Slow start suggests that the sender set the congestion window to 1 and then for each ACK received it increase the CWD by 1. So in the first round trip time (RTT) we send 1 packet, in the second we send 2 and in the third we send 4. Thus we increase exponentially until we lose a packet which is a sign of congestion. When we encounter congestion we decrease our sending rate and we reduce congestion window to one. And start over again.

The important thing is that Tahoe detects packet losses by timeouts. In usual implementations, repeated interrupts are expensive so we have coarse grain time-outs which occasionally checks for time outs. Thus it might be some time before we notice a packet loss and then re-transmit that packet.

Problem – The problem with Tahoe is that it takes a complete timeout interval to detect a packet loss and in fact, in most implementations it takes even longer because of the coarse grain timeout. Also since it doesn't send immediate ACK's, it sends cumulative acknowledgements, therefore it follows a 'go back n' approach. Thus every time a packet is lost it waits for a timeout and the pipeline is emptied. This offers a major cost in high bandwidth delay product links.

Congestion Avoidance – For congestion avoidance Tahoe uses 'Additive Increase Multiplicative Decrease'. A packet loss is taken as a sign of congestion and Tahoe saves the half of the current window as a threshold value. It then set CWD to one and starts slow start until it reaches the threshold value. After that it increments linearly until it encounters a packet loss. Thus it increases its window slowly as it approaches the bandwidth capacity.

Q.35. Discuss the TCP Reno congestion control and avoidance mechanism.

Ans. Reno retains the basic principle of Tahoe. Reno requires that receive immediate acknowledgement whenever a segment is received. The logic behind this is that whenever we receive a duplicate acknowledgement, then his duplicate acknowledgement could have been received if the segment in sequence expected, has been delayed in the network and the segments reached there out of order or else that the packet is lost. If we receive a number of duplicate acknowledgements then that means that sufficient time has passed and even if the segment had taken a longer path, it should have gotten to the receiver by now. There is a very high probability that it was lost. So Reno suggest an algorithm called 'Fast Re-Transmit'. Whenever we receive 3 duplicate ACK's we take it as a sign that the segment was lost, so we re-transmit the segment without waiting for timeout. Thus we manage to transmit the segment with the pipe almost full.

Another modification that Reno makes is in that after a packet loss does not reduce the congestion window to 1. Since this empties the pipe, it enters into a algorithm which we call 'Fast-Re-transmit'. The basic algorithm is presented as under –

(i) Each time we receive 3 duplicate ACK's we take that to mean the segment was lost and we re-transmit the segment immediately and enter 'Fast-Recovery'.

(ii) Set SS_{thresh} to half the current window size and also set CWD to the same value.

(iii) For each duplicate ACK receive increase CWD by one. If CWD increase CWD is greater than the amount of data in the pipe then transmit new segment else wait. If there are 'w' segments in the window and one is lost then we will receive (w-1) duplicate ACK's. Since CWD is reduced to SS_{thresh} therefore half a window of data is acknowledged before we can send a new segment. Once we retransmit a segment, we would have to wait for at least one RTT before we would receive a fresh acknowledgement. Whenever we receive a fresh ACK we reduce the $CWND$ to SS_{thresh} . If had previously received (w-1) duplicate ACK's then at the point we should have exactly w segments in the pipe which is equal to what we set the $CWND$ to be at the start of fast recovery. Thus we don't empty the pipe, we just reduce the flow and continue with congestion avoidance phase of Tahoe after that.

Problems – Reno perform very well over TCP when the packet loss is small. But when we have multiple packet losses in one window then Reno doesn't perform too well and its performance is almost the same as Tahoe under conditions of high packet loss. The reason is that it can only detect single packet losses. If there is multiple packet drop then the first info about

the packet loss comes when we receive the duplicate ACK's. But the information about the second packet which was lost will come only after the ACK for the re-transmitted first segment reaches the sender after one RTT.

Also it is possible that the CWD is reduced twice for packet losses which occurred in one window. Suppose we send packet 1, 2, 3, 4, 5, 6, 7, 8, 9 in that order. Suppose packets 1, and 2 are lost. The ACK's generated by 2, 4, 5 will cause the re-transmission of 1 and the CWD is reduced to 7. Then when we receive ACK for 6, 7, 8, 9 our CWD is sufficiently large to allow us to send 10, 11. When the re-transmitted segment 1 reaches the receiver we get a fresh ACK and we exit fast-recovery and set CWD to 4. Then we get two more ACK's for 2 (due to 10, 11) so once again we enter fast-retransmit and re-transmit 2 and then enter fast recovery. Thus when we exit fast recovery for the second time our window size is set to 2. Thus we reduced our window size twice for packets lost in one window.

Another problem is that if the window is very small when the loss occurs then we would never receive enough duplicate acknowledgements for a fast-retransmit and we would have to wait for a coarse grained timeout. Thus it cannot effectively detect multiple packet losses.

Q.36. Discuss the new Reno congestion control and avoidance mechanism.

Ans. New Reno is a slight modification over TCP-Reno. It is able to detect multiple packet losses and thus is much more efficient than Reno in the event of multiple packet losses.

Like Reno, New-Reno also enters into fast-retransmit when it receives multiple duplicate packets, however it differs from Reno in that it doesn't exit fast-recovery until all the data which was outstanding at the time it entered fast-recovery is acknowledged. Thus it overcomes the problem faced by Reno of reducing the CWD multiples times.

The fast-transmit phase is the same as in Reno. The difference in the fast-recovery phase which allows for multiple re-transmissions in new-Reno. Whenever new-Reno enters fast-recovery it notes the maximum segment which is outstanding. The fast-recovery phase proceeds as in Reno, however when a fresh ACK is received then there are two cases –

If it ACK's all the segments which were outstanding when we entered fast-recovery then it exits fast recovery and sets CWD to ss_{thresh} and continues congestion avoidance like Tahoe.

If the ACK is a partial ACK then it deduces that the next segment in line was lost and it re-transmits that segment and sets the number of duplicate ACK's received to zero.

It exits fast recovery when all the data in the window is acknowledged.

Problems – New-Reno suffers from the fact that it takes one RTT to detect each packet loss. When the ACK for the first retransmitted segment is received only then can we deduce which other segment was lost.

Q.37. Explain the SACK congestion control and avoidance mechanism.

Ans. TCP with 'Selective Acknowledgements' is an extension of TCP Reno and it works around the problems faced by TCP Reno and TCP New-Reno, namely detection of multiple lost packets, and re-transmission of more than one lost packet per RTT.

SACK retains the slow-start and fast-retransmit parts of Reno. It also has the coarse grained timeout of Tahoe to fall back on, in case a packet loss is not detected by the modified algorithm.

SACK TCP requires that segments not be acknowledged cumulatively but should be acknowledged selectively. Thus each ACK has a block which describes which segments are being acknowledged. Thus the sender has a picture of which segments have been acknowledged and which are still outstanding. Whenever the sender enters fast recovery, it initializes a variable pipe which is an estimate of how much data is outstanding in the network, and it also sets CWND to half the current size. Every time it receives an ACK it reduces the pipe by 1 and every time it retransmits a segment it increments it by 1. Whenever the pipe goes smaller than the CWD window it checks which segments are unreceived and sends them. If there are no such segments outstanding then it sends a new packet. Thus more than one lost segment can be sent in one RTT.

Problems – The biggest problem with SACK is that currently selective acknowledgements are not provided by the receiver to implement SACK we will need to implement selective acknowledgement which is not a very easy task.

Q.38. Discuss in detail Vegas congestion control and avoidance mechanism.

Ans. Vegas is a TCP implementation which is a modification of Reno. It builds on the fact that proactive measures to encounter congestion are much more efficient than reactive ones. It tried to get around the problem of coarse grain timeouts by suggesting an algorithm which checks for timeouts at a very efficient schedule. Also it overcomes the problem of requiring enough duplicate acknowledgements to detect a packet loss, and it also suggests a modified slow start algorithm which prevents it from congesting the network. It does not depend solely on packet loss as a sign of congestion. It detects congestion before the packet losses occur. However it still retains the other mechanism of Reno and Tahoe, and a packet loss can still be detected by the coarse grain timeout of the other mechanisms fail.

The three major changes induced by Vegas are –

New Re-Transmission Mechanism – Vegas extends on the re-transmission mechanism of Reno: It keeps track of when each segment was sent and it also calculates an estimate of the RTT by keeping track of how long it takes for the acknowledgement to get back. Whenever a duplicate acknowledgement is received it checks to see if the (current time-segment transmission time) > RTT estimate; if it is then it immediately retransmits the segment without waiting for 3 duplicate acknowledgements or a coarse timeout. Thus it gets around the problem faced by Reno of not being able to detect lost packets when it had a small window and it didn't receive enough duplicate Ack's.

To catch any other segments that may have been lost prior to the re-transmission, when a non duplicate acknowledgement is received, if it is the first or second one after a fresh acknowledgement then it again checks the timeout values and if the segment time since it was sent exceeds the timeout value then it re-transmits the segment without waiting for a duplicate acknowledgement. Thus in this way Vegas can detect multiple packet losses.

Also it only reduces its window if the re-transmitted segment was sent after the last decrease. Thus it also overcomes Reno's shortcoming of reducing the congestion window multiple times when multiple packets are lost.

Congestion Avoidance – TCP Vegas is different from all the other implementations in its behaviour during congestion avoidance. It does not use the loss of segment to signal that there is congestion. It determines congestion by a decrease in sending rate as compared to the expected rate, as a result of large queues building up in the routers. Thus whenever the calculated rate is too far away from the expected rate it increases transmissions to make use of the available bandwidth, whenever the calculated rate comes too close to the expected value it decreases its transmission to prevent over saturating the bandwidth. Thus Vegas combats congestion quite effectively and doesn't waste bandwidth by transmitting at too high a data rate and creating congestion and then cutting back, which the other algorithms do.

Modified Slow-start – TCP Vegas differs from the other algorithms during its slow-start phase. The reason for this modification is that when a connection first starts it has no idea of the available bandwidth and it is possible that during exponential increase it over shoots the bandwidth by a big amount and thus induces congestion. To this end Vegas increases exponentially only every other RTT, between that it calculates the actual sending throughput to the expected and when the difference goes above a certain threshold it exits slow start and enters the congestion avoidance phase.

UDP – MESSAGE ENCAPSULATION, FORMAT AND PSEUDO HEADER

Q.39. Write short note on UDP.

Ans. UDP (User Datagram Protocol) is another transport layer protocol that is a part of TCP/IP suite. It is an unreliable, connectionless protocol that does not guarantee delivery and duplicate protection. UDP is basically just IP with a short header added. It provides non-sequential transport functionality when reliability and security are less important than size and speed.

UDP provides a way for applications to send encapsulated raw IP datagrams without establishing any connection. UDP is more appropriate than TCP in some contexts; like it is more robust at lower layers. UDP has very little to do. Mainly, it adds a port addressing capability to IP. Since its header is shorter, it involves less overheads.

UDP is used by the applications that do not want TCP's sequencing and flow control and wish to provide their own. It is widely used for one-shot, client-server type request-reply queries and applications in which prompt delivery is more important than accurate delivery, such as transmitting speech or video. Thus, there is a place at the transport layer for both connection-oriented and connectionless type of service. UDP is described in RFC 768.

The relation of IP, TCP and UDP is shown in fig. 2.43.

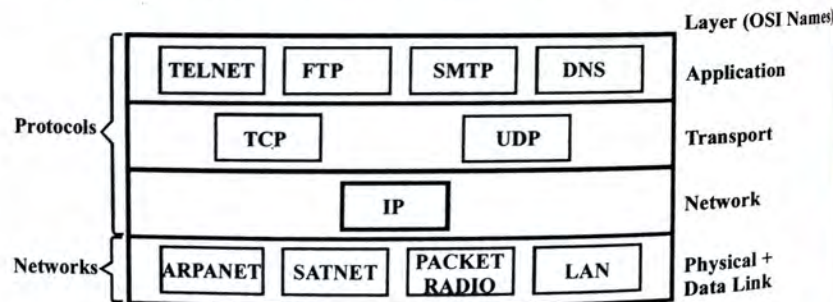


Fig. 2.43 Protocols and Networks in TCP/IP Model

Q.40. What does UDP provide that is not provided by IP ?

Ans. UDP provides two services not provided by the IP layer. It provides port number to help distinguish different user requests and, optionally, a checksum capability to verify that the data arrived intact.

Q.41. What are the advantages of UDP ?

Ans. The advantages of UDP are as follows –

- (i) **Simplicity** – For the network not for the programmer.
- (ii) **Does not Include the Overhead Needed to Detect Reliability and Maintain Connection Oriented Semantics** – UDP packets require considerably less processing at the transmitting and receiving hosts.
- (iii) **Less Heavy to Manage for the OS** – Many operating systems impose limits on how many simultaneous TCP/IP connections they can support.

Q.42. What are the uses of UDP ?

Ans. Some of the uses of the UDP protocol are –

- (i) UDP is suitable for a process that needs simple request-response communication with little concern for flow and error control. Generally, it is not used for a process that requires to send bulk data, such as FTP.
- (ii) UDP is suitable for a process with internal flow and error control mechanisms. For example, TFTP (Trivial File Transfer Protocol) process includes flow and error control. It can easily use UDP.
- (iii) UDP is used for management processes such as SNMP.
- (iv) UDP is suitable transport protocol for multicasting and broadcasting. These facilities are built in the UDP software but not in the TCP software.
- (v) UDP is used for some route updating protocols such as routing information protocol (RIP).

Q.43. Explain UDP encapsulation and decapsulation.

Ans. The UDP protocol encapsulates and decapsulates messages to be sent from one process to another (see fig. 2.44).

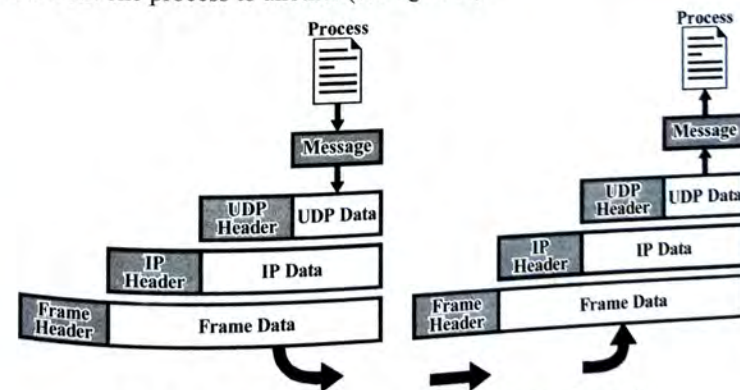


Fig. 2.44 Encapsulation and Decapsulation Operations

Encapsulation – When a process has a message to send through UDP, it passes the message to UDP along with a pair of socket addresses and the length of data. UDP receives the data and adds the UDP header. UDP then passes the user datagram to the IP with the socket addresses. IP adds its own header, using the value 17 in the protocol field, indicating that the data has come from the UDP protocol. The IP datagram is then passed to the data link layer. The data link layer receives the IP datagram, adds its own header and passes it to the physical layer. The physical layer encodes the bits into electrical or optical signals and sends it to the remote machine.

Decapsulation – When the message arrives at the destination host, the physical layer decodes the signals into bits and passes it to the data link layer. The data link layer uses the header to check the data. If there is no error, the header and the trailer are dropped and the user datagram is passed to the IP. The IP software does its checking. If there is no error, the header is dropped and the user datagram is passed to the UDP with the sender and receiver IP addresses. UDP uses the checksum to check the entire user datagram. If there is no error, the header is dropped and the application data along with the sender socket address is passed to the process. The sender socket address is passed to the process in case it needs to respond to the message received.

Q.44. Explain the frame format of a user datagram protocol (UDP).

Ans. UDP packets, called user datagrams, have a fixed-size header of 8 bytes. Fig. 2.45 shows the format of a user datagram.

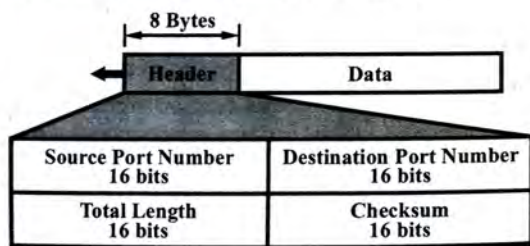


Fig. 2.45 User Datagram Format

The fields are as follows –

(i) **Source Port Number** – This is the port number used by the process running on the source host. It is 16 bits long, which means that the port number can range from 0 to 65,535. If the source host is the client, the port number, in most cases, is an ephemeral port number requested by the process and chosen by the UDP software running on the source host. If the source host is the server (a server sending a response) the port number, in most cases, is a well-known port number.

(ii) **Destination Port Number** – This is the port number used by the process running on the destination host. It is also 16 bits long. If the destination host is the server (a client sending a request), the port number, in most cases, is a well-known port number. If the destination host is the client (a server sending a response), the port number, in most cases is an ephemeral port number. In this case, the server copies the ephemeral port number it has received in the request packet.

(iii) **Length** – This is a 16-bit field that defines the total length of the user datagram, header plus data. The 16 bits can define a total length of 0 to 65,535 bytes. However, the total length needs to be much less because a UDP user datagram is stored in an IP datagram with a total length of 65,535 bytes.

The length field in a UDP user datagram is actually not necessary. A user datagram is encapsulated in an IP datagram. There is a field in the IP datagram that defines the total length. There is another field in the IP datagram that defines the length of the header. So if we subtract the value of the second field from the first, we can deduce the length of a UDP datagram that is encapsulated in an IP datagram.

UDP length = IP length – IP header's length

(iv) **Checksum** – This field is used to detect errors over the entire user datagram.

Q.45. Give the format of UDP datagram and explain the semantics of every field. How is the checksum in the header computed?

Ans. UDP Datagram Format – Refer the ans. of Q.44.

Computation of Checksum – The UDP checksum calculation is different from the one for IP and ICMP. Here the checksum includes three sections – a pseudoheader, the UDP header, and the data coming from the application layer.

The pseudoheader is the part of the header of the IP packet in which the user datagram is to be encapsulated with some fields filled with 0s (see fig. 2.46).

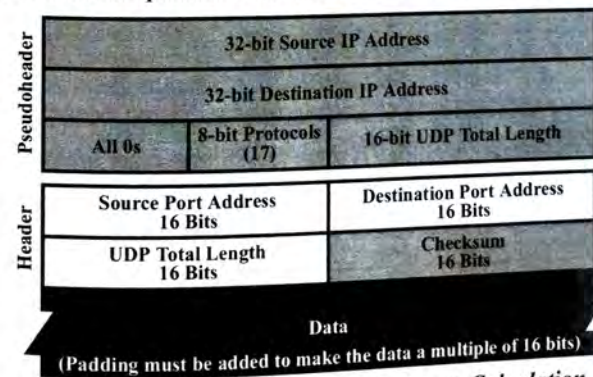


Fig. 2.46 Pseudoheader for Checksum Calculation

If the checksum does not include the pseudoheader, a user datagram may arrive safe and sound. However, if the IP header is corrupted, it may be delivered to the wrong host.

The protocol field is added to ensure that the packet belongs to UDP, and not to other transport-layer protocols. The value of the protocol field for UDP is 17. If this value is changed during transmission, the checksum calculation at the receiver will detect it and UDP drops the packet. It is not delivered to the wrong protocol.



UNIT

3

WIRELESS LAN

TRANSMISSION MEDIUM FOR WLANS, MAC PROBLEMS, HIDDEN AND EXPOSED TERMINALS, NEAR AND FAR TERMINALS, INFRASTRUCTURE AND AD-HOC NETWORKS

Q.1. What do you understand by wireless LAN ? Write down some important requirements for wireless LANs. (R.G.P.V., Nov. 2018)

Ans. A wireless LAN is one that makes use of a wireless transmission medium. WLAN is a fast growing market introducing the flexibility of wireless access into office, home, or production environments. WLANs are typically restricted in their diameter to buildings, a campus single room etc., and are operated by individuals, not by large-scale providers. The main goal of WLANs is to replace office cabling and, additionally to introduce a higher flexibility for ad-hoc communication. Until relatively recently, wireless LANs were little used. The reasons for this included high prices, low data rates, occupational safety concerns, and licensing requirements. As these problems have been addressed, the popularity of wireless LANs has grown rapidly.

A wireless LAN must satisfy the some type of requirements typical of any LAN, including high capacity, ability to cover short distances, full connectivity among attached stations, and broadcast capability. Furthermore, there are a number of requirements specific to the wireless LAN environment. Following are the main requirements for wireless LANs –

(i) **Throughput** – To maximize capacity, the medium access control protocol should make as efficient use as possible of the wireless medium.

(ii) **Number of Nodes** – Wireless LANs may require to support hundreds of nodes across multiple cells.

(iii) **Connection to Backbone LAN** – Generally, interconnection with stations on a wired backbone LAN is needed. This is easily accomplished for infrastructure wireless LANs through the use of control modules that connect to both types of LANs. There may also require to be accommodation for mobile users and ad-hoc wireless networks.

(iv) **Service Area** – For a wireless LAN, a typical coverage area has a diameter of 100 to 300 m.

(v) **Battery Power Consumption** – Mobile workers use battery powered workstations that require to have a long battery life when used with wireless adapters. This suggests that a MAC protocol that needs mobile nodes to monitor access points constantly or engage in frequent handshakes with a base station is not suitable. Typical wireless LAN implementations have features to decrease power consumption while not using the network, like a sleep mode.

(vi) **Transmission Robustness and Security** – Unless appropriately designed, a wireless LAN may be interference prone and easily eavesdropped. The design of a wireless LAN must allow reliable transmission even in noisy environment and should provide some level of security from eavesdropping.

(vii) **Collocated Network Operation** – When wireless LANs become more popular, it is quite likely for two or more wireless LANs to operate in the same area or in some area where interference between the LANs is possible. This type of interference may thwart the normal operation of a MAC algorithm and may permit unauthorized access to a particular LAN.

(viii) **License-free Operation** – Users would prefer to buy and operate wireless LAN products without having to secure a license for the frequency band employed by the LAN.

(ix) **Handoff/Roaming** – In wireless LAN, the MAC protocol used should enable mobile stations to move from one cell to another.

(x) **Dynamic Configuration** – MAC addressing and network management aspects of the LAN should allow dynamic and automated addition, deletion, and relocation of end systems without disruption to other users.

Q.2. Write and explain the four major challenges for implementing wireless LAN. (R.G.P.V., Dec. 2015)

Ans. Several important challenges facing the WLAN industry –

(i) **Complexity and Cost** – The alternatives for implementing WLANs such as IR (infrared), spread spectrum, or traditional radios are far more complex and diversified than the wired LANs.

(ii) **Bandwidth** – Data rate limitations of the wireless medium are more serious than those of wired media.

(iii) **Coverage** – The coverage of a WLAN operating within a building is less than that of a single cable or even TP-based LANs.

(iv) **Interference** – WLANs are subject to interference from other overlaid WLANs or other users operating in the same frequency bands.

(v) **Frequency Administration** – Radio-based WLANs are subject to expensive and untimely frequency regulations.

Q.3. Write advantages and disadvantages of WLANs.

Or

(R.G.P.V., Dec. 2010)

Write WLAN advantages and disadvantages.

(R.G.P.V., Dec. 2015)

Ans. Advantages of WLANs – Advantages of wireless LANs are as follows –

(i) Cables not only restrict users but also designers of small PDAs, notepad etc. Only wireless networks allow for the design of small, independent devices.

(ii) Nodes can communicate without restriction. Senders and receivers can communicate from anywhere.

(iii) Wireless networks can survive disasters, like, earthquakes.

(iv) Wireless networks also make it easier for any office – in either an old or new building – to be modified with the new cubicles or furniture. No longer does the design for a remodeled office.

(v) Network cable failures are perhaps the most common source of network problems. Moisture from a leak during a thunderstorm can erode metallic conductors. A user who moves a wired computer may break the network connection. A WLAN eliminates these types of cable failures and increases the overall reliability of the network.

Disadvantages of Wireless LANs –

(i) **Restrictions** – All wireless products have to comply with national regulations. Several government and non government institutions worldwide regulate the operation and restrict frequencies to minimize interference. Consequently, it takes a very long time to establish global solutions like, e.g., IMT-2000. WLANs are limited to low-power senders and certain licence-free frequency bands.

(ii) **Cost** – While, e.g., high-speed Ethernet adapters are in the range of some 10\$, wireless LAN adapters, e.g., as PC-card, still cost some 100\$.

(iii) WLANs typically offer lower quality than their wired counterparts. The main reasons for this drawback are the lower bandwidth due to limitations in radio transmission (e.g., only 1-10 Mbit/s), higher error rates due to interference (e.g., 10^{-4} instead of 10^{-10} for fibre optics).

(iv) **Safety and Security** – Using radio waves for data transmission might interfere with other high-tech equipment in, e.g., hospitals.

Q.4. What do you understand by wireless LAN? Write advantages and disadvantages of WLAN. (R.G.P.V., Dec. 2017)

Ans. Refer the ans. of Q.1 and Q.3.

Q.5. How authentication is possible in wireless LAN? List and discuss the possible attacks on such network. (R.G.P.V., Nov. 2019)

Ans. There are three main methods of authentication that are used on today's wireless LANs –

- (i) Open authentication
- (ii) Shared authentication
- (iii) EAP (extensible authentication protocol) authentication.

The open authentication method is the simplest of the methods used and only requires that the end device be aware of the service-set identifier (SSID) used on the network, as long as the SSID is known then the device will be allowed onto the network. The problem with this method is that the SSID is typically broadcast and if it is not, it can be easy to figure out with passive capturing techniques.

The shared authentication method is commonly used on individual and small business wireless LAN implementations; this method uses a shared key (Pre-shared Key-PSK) that is given to both sides of the connection; if they match then the device is allowed onto the network.

The third method uses the extensible authentication protocol (EAP) and is the most common method used by enterprises. The EAP method utilizes an authentication server that is queried for authentication using a variety of credential options.

The most common types of attack are as follows –

(i) Probing and Discovery Tools – A host of tools have emerged that take advantage of the fact that 802.11 infrastructures rely on network broadcasts to communicate with wireless clients. With these probing and discovery tools, unscrupulous individuals can easily locate, and take advantage of, wireless networks that lack strong security safeguards. One of the most common of these tools is Netstumbler, a Windows-based program that uses active scanning to detect low security access points. Once the access point is detected a number of exploits can be mounted against the network.

(ii) MAC Identify Spoof Attacks – In an 802.11 WLAN, MAC addresses are openly broadcasted over the air. The security implication is that potential attackers can sniff the air looking for valid MAC addresses associated with authorized WLAN users, access points and even wired infrastructure components, such as switches and routers. Once detected, programs exist to spoof these addresses, whereby intruders can masquerade as a valid WLAN client or access point. Naturally, this can compromise a WLAN if MAC authentication is the only security scheme employed.

(iii) Denial of Service Attacks – Because WLANs broadcast over the unlicensed ISM & U-NII public bands with a limited number of available

channels, RF interference is a common problem. As interference increases, signal quality and network availability decreases. Malicious individuals can use this to their advantage, debilitating WLAN performance. Common ways of doing this include RF frequency jamming and exploits such as Airjack and void11, which flood the WLAN.

(iv) Man in the Middle Attacks – A man in the middle attack results from the interception and possible modification of traffic passing between two communicating parties, such as a wireless client and AP. Man in the middle attacks succeed if the systems can't distinguish communications with an intended recipient from those with the intervening attacker.

(v) Static WEP Cracking Programs – Soon after it was first introduced, the Wired Equivalency Protocol (WEP) was broken due to the fact that WEP uses static keys which can be easily cracked. While these deficiencies were soon corrected with the WiFi Protected Access (WPA) protocol and 802.11i (WPA2), both of which leverage dynamic keys, many WLANs continue to use WEP-based security. As a result, they are still vulnerable to WEP cracking programs.

(vi) Rogue Access Point Attack Programs – A number of "rogue access point attack programs" exist which allow attackers to perform a number of stealth attacks by posing as host access points on a WLAN network.

(vii) Misconfigured Clients – Due to the nature of the 802.11 specification enterprise WLANs are vulnerable to security risks when new hosts or clients enter the network and when ad-hoc networking is allowed. A wired host with an enabled WLAN adapter, for example, could unwittingly connect to an unknown WLAN. An attacker would then be able to compromise the host machine via the open WLAN adapter through routing features on Linux and Windows and mount an attack against the wired connection. Similarly, the overflow of RF signals means that accidental connections can occur with neighbouring WLANs, which can compromise the security of trusted networks.

Q.6. What are the basic difference between wireless WANs and wireless LANs and what are the common features? (R.G.P.V., June 2005)

Ans. Differences between Wireless WANs and Wireless LANs –

(i) Global Operation – WLAN products should sell in all countries so, national and international frequency regulations have to be considered. In contrast to the infrastructure of wireless WANs, LAN equipment may be carried from one country into another – the operation should still be legal in this case.

(ii) Low Power – Devices communicating via a WLAN are typically also wireless devices running on battery power. The LAN design should take

this into account and implement special power-saving modes and power management functions. Wireless communication with devices plugged into a power outlet is only useful in some cases.

(iii) Robust Transmission Technology – Compared to their wired counterparts, WLANs operate under difficult conditions. If they use radio transmission, many other electrical devices can interface with them. WLAN transceivers cannot be adjusted for perfect transmission in a standard office or production environment. Antennas are typically omni-directional, not directed. Senders and receivers may move.

(iv) Simplified Spontaneous Cooperation – To be useful in practice WLANs should not require complicated setup routines but should operate spontaneously after power-up. These LANs would not be useful for supporting e.g., ad-hoc meetings.

(v) Easy to Use – In contrast to huge and complex wireless WANs, wireless LANs are made for simple use. They should not require complex management but rather work on plug-and-play basis.

(vi) Protection of Investment – A lot of money has already been invested into wired LANs. The new WLANs should protect this investment by being interoperable with the existing networks.

(vii) Safety and Security – Wireless LANs should be safe to operate, especially regarding low radiation if used, e.g., in hospitals. Users cannot keep safety distances to antennas. The equipment has to be safe for pacemakers, too. Users should not be able to read personal data during transmissions, i.e., encryption mechanisms should be integrated. The networks should also take into account user privacy, i.e., it should not be possible to collect roaming profiles for tracking persons if they do not agree.

(viii) Transparency for Applications – Existing applications should continue to run over WLANs, the only difference being higher delay and lower bandwidth. The fact of wireless access and mobility should be hidden if it is not relevant, but the network should also support location aware applications, e.g., by providing location information.

Common Features –

- (i) Freedom to access shared information from anywhere in a building, company or organization without being tied to a cable connection.
- (ii) Enhances productivity and efficiency by placing the internet, email, corporate data and applications constantly at the fingertips.
- (iii) Access to real time information in meeting rooms to help with decision making.
- (iv) Notebook users never lose connection, even when moving from office-to-office.

Q.7. Explain about wireless transmission.

Ans. Wireless transmission is a form of unguided media. Wireless communication involves no physical link established between two or more devices, communicating wirelessly. Wireless signals are spread over in the air and are received and interpreted by appropriate antennas.

When an antenna is attached to electrical circuit of a computer or wireless device, it converts the digital data into wireless signals and spread all over within its frequency range. The receptor on the other end receives these signals and converts them back to digital data.

A little part of electromagnetic spectrum can be used for wireless transmission.

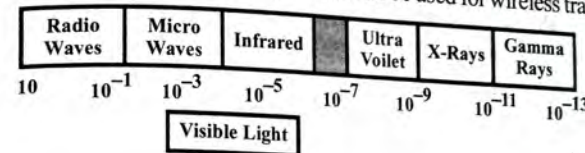


Fig. 3.1

Radio Wave Transmission – Radio waves are easy to generate, can travel long distances, and can penetrate buildings easily, so they are widely used for communication both indoors and outdoors. Radio waves also are omnidirectional, meaning that they travel in all directions from the source, so the transmitter and receiver do not have to be carefully aligned physically.

The properties of radio waves are frequency dependent. At low frequencies, radio waves pass through obstacles well, but the power falls off sharply with distance from the source roughly as $1/r^2$ in air. At high frequencies, radio waves tend to travel in straight lines and bounce off obstacles. They are also absorbed by rain. At all frequencies, radio waves are subject to interference from motors and other electrical equipment.

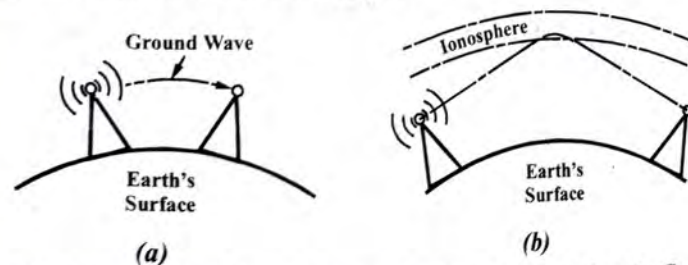


Fig. 3.2 (a) In the VLF, LF and MF Bands, Radio Waves follow the Curvature of the Earth (b) In the HF Band they Bounce off the Ionosphere

Due to radio's ability to travel long distances, interference between users is a problem. In the VLF, LF and MF bands, radio waves follow the ground. In the HF and VHF bands, the ground waves tend to be absorbed by the earth.

However, the waves that reach the ionosphere, a layer of charged particles circling the earth at a height of about 100 to 500 km, are refracted by it and sent back to earth. As shown in figs. 3.2 (a) and (b).

Microwave Transmission – This is a popular way of transmitting data since it does not incur the expense of laying cables. Microwave systems use very high frequency radio signals to transmit data through space. However, at microwave frequencies the electromagnetic waves cannot bend or pass obstacles like hill. Hence it is necessary that microwave transmission is in a line-of-sight. In other words, the transmitter and receiver of a microwave system, which are mounted on very high towers, should be in a line-of-sight. This may not be possible for very long distance transmission. Moreover, the signals become weaker after travelling a certain distance and require power amplification.

In order to overcome the problem of line-of-sight and power amplification of weak signals, microwave systems use repeaters at intervals of about 25 to 30 kms in between the transmitting and receiving stations as shown in fig. 3.3. The first repeater is placed in line-of-sight of the transmitting station and the last repeater is placed in line-of-sight of the receiving station. Two consecutive repeaters are also placed in line-of-sight of each other. The data signals are received, amplified, and retransmitted by each of these stations.

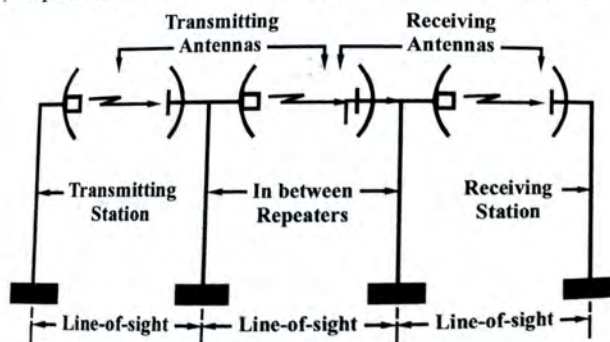


Fig. 3.3 Illustrating Microwave Communication from the One Point to Another

Infrared Transmission – For short distance communication, the waves with frequency range from 300 GHz to 400 THz i.e., wavelengths from 1 mm to 770 nm can be used. High frequency infrared waves cannot penetrate walls. This property prevents interference between one system and another, a short-range communication system in one room cannot be affected by another system in the other room. For example, if we use our infrared remote control we do not interfere with the use of the remote by our neighbours. Although for long distance communication, the same property makes infrared signal useless. Furthermore, we cannot use infrared waves outside a building as solar rays containing infrared waves can interfere with the communication.

Application of Infrared Transmission – There is an excellent potential for data transmission in the infrared band which is almost 400 THz wide. This wide bandwidth can be used for sending digital data at very high data rate. For sponsoring the use of infrared waves, the infrared data association (IrDA) has established standards to use these signals for communication between devices like PCs, printers, keyboards and mice. As an example, few manufacturers offer a special port referred as the IrDA port which permits using a wireless keyboard for communicating with a PC. Earlier, the standard defines a data rate of 75 kbps upto 8 metre distance. However, a 4 Mbps data rate is defined by the latest standard.

Light Transmission – Highest most electromagnetic spectrum which can be used for data transmission is light or optical signaling. This is achieved by means of LASER.

Because of frequency light uses, it tends to travel strictly in straight line. Hence the sender and receiver must be in the line-of-sight. Because laser transmission is unidirectional, at both ends of communication the laser and the photo-detector needs to be installed. Laser beam is generally 1 mm wide hence it is a work of precision to align two far receptors each pointing to lasers source.

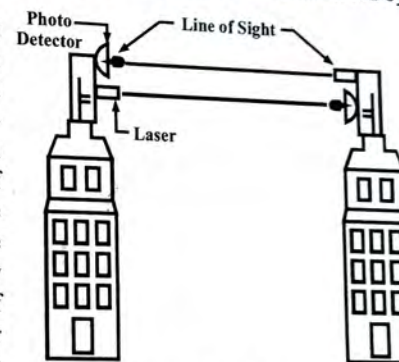


Fig. 3.4

Laser works as transmitter and photo-detectors works as receiver.

Lasers cannot penetrate obstacles such as walls, rain, and thick fog. Additionally laser beam is distorted by wind, atmosphere temperature, or variation in temperature in the path.

Laser is safe for data transmission as it is very difficult to tap 1 mm wide laser without interrupting the communication channel.

Q.8. Write a short note on media access control.

(R.G.P.V., Dec. 2005, 2007, June 2010)

Ans. Medium access control comprises all mechanisms which regulate user access to a medium using SDM, TDM, FDM or CDM. Medium access control (MAC) algorithms are specifically designed to the wireless domain. Hence, MAC is similar to traffic regulations in the highway/multiplexing. As an example, the fact that several vehicles use the same street crossing in TDM needs rules to avoid collisions; one mechanism to enforce these rules is traffic lights. MAC belongs to layer 2 of the ISO/OSI reference model, the data link

control layer (DLC). Layer 2 is subdivided into the logical link control (LLC), layer 2b, and the MAC, layer 2a. The function of the DLC is to establish a reliable point-to-point or point-to-multipoint connection between different devices over a wired or wireless medium. Several MAC mechanisms are presented for the multiplexing schemes, in which SDM and FDM are typically used in a rather fixed manner, i.e., a certain space or frequency or (frequency hopping pattern) is allocated for a longer period of time, while TDM can be used in a very flexible way, as tuning into a certain frequency does not present a problem, but time can be assigned on demand and in a distributed manner. Well known algorithms are Aloha (in several versions) different reservations schemes, or simple polling. A MAC scheme using CDM has to assign certain codes to permit the separation of different users in code space.

Q.9. Explain the following –

- (i) Hidden and exposed terminals problems
- (ii) Near and far terminals problems.

(R.G.P.V., Dec. 2011, 2017)

Ans. The problems associated with 'hidden and exposed terminals' and 'near and far terminals' are given below –

(i) **Hidden and Exposed Terminals Problems** – Fig. 3.5 shows the case with three mobile phones. The transmission range of A reaches B, but not C. The transmission range of C reaches B, but not A. At last, the transmission range of B reaches A and C, i.e., A cannot detect C and vice-versa.

A begins sending to B, C does not receive this transmission. In addition, C desires to send something to B and senses the medium. The medium appears to be free, the carrier sense fails C also begins sending causing a collision at B. However, A cannot detect this collision at B and continues with its transmission. A is hidden for C and vice-versa.

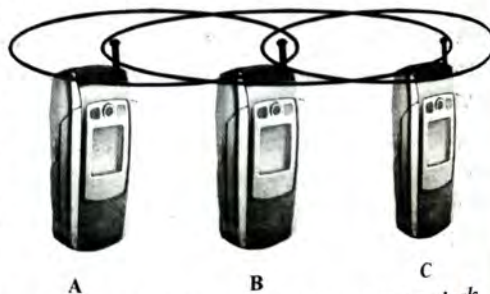


Fig. 3.5 Hidden and Exposed Terminals

While hidden terminals can cause collisions, the next effect only causes unnecessary delay. Now, we take the case in which B sends something to A and C wants to send data to some other mobile phone outside the interference ranges of A and B. C senses the carrier and detects that the carrier is not free (B's signal). C postpones its transmission until it detects the medium as being idle again. But as A is outside the interference range of C, waiting is not necessary. Causing a collision at B does not matter since the collision is too weak to propagate to A. In this situation, C is exposed to B.

(ii) **Near and Far Terminals Problems** – Fig. 3.6 shows the case, in which A and B, both, are sending with the same transmission power. Since the signal strength decreases proportionally to the square of the distance, B's signal drowns out A's signal. Due to this, C cannot receive A's transmission.

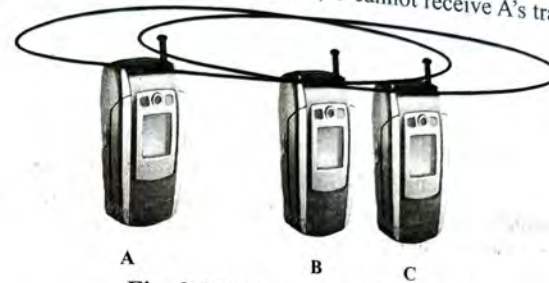


Fig. 3.6 Near and Far Terminals

Now, let us consider the case of C as being an arbiter for sending rights (e.g., C acts as a base station coordinating media access). In such a situation, terminal B would already drown out terminal A on the physical layer. In return, C would have no opportunity of applying a fair scheme as it would only hear B.

In the wireless networks using CDM, near/far effect is a serious problem. At the receiver, all signals should arrive with more or less the same strength. Otherwise, a person standing closer to somebody could always speak louder than a person further away. Even if the senders were separated by code, the closest one would simply drown out the others. Accurate power control is required to receive all senders with the same strength at a receiver.

Q.10. Why a specialized MAC is used in wireless network? Explain the problems associated with 'hidden and exposed terminals' and 'near and far terminals'.

(R.G.P.V., Nov. 2018)

Ans. In the wireless, it is not possible to use elaborated MAC schemes from wired networks, for example, CSMA/CD as used in the original specification of IEEE 802.3 networks (aka Ethernet). Let us consider carrier sense multiple access with collision detection (CSMA/CD) which works as follows. A sender senses the medium to see if it is free. In case the medium is not free, the sender waits until it is free. When the medium is free, the sender begins transmitting data and continues to listen into the medium. When sender detects a collision while sending, it stops at once and sends a jamming signal.

This scheme fails in wireless networks because CSMA/CD is not actually interested in collisions at the sender, but rather in those at the receiver. The signal should reach the receiver without collisions. But, the sender is the one detecting collisions. When we are using a wire, this is not a problem as more or less the same signal strength can be assumed all over the wire if the length of the wire stays within certain often standardized limits. When a collision

takes place somewhere in the wire, everybody will notice it. It does not matter if a sender listens into the medium to detect a collision at its own location while in reality is waiting to detect a possible collision at the receiver.

In wireless networks, the situation is different. The strength of the signal weakens proportionally to the square of the distance to the sender. Obstacles attenuate the signal even further. Now, the sender can apply carrier sense and detect an idle medium. The sender begins sending – but a collision occurs at the receiver owing to a second sender. This is hidden terminal problem. The same can occur to the collision detection. The sender detects no collision and assumes that the data has been sent without errors, but a collision might really have destroyed the data at the receiver. In wireless scenarios, collision detection is very difficult, as the transmission power in the area of the transmitting antenna is several magnitudes higher than the receiving power. Therefore, this very common MAC scheme from wired network fails to a wireless scenario.

Also refer the ans. of Q.9.

Q.11. Illustrate the problem of near and far terminals in medium access control.

(R.G.P.V., Nov. 2019)

Ans. Refer the ans. of Q.9 (ii).

Q.12. What is the main physical reason for the failure of many MAC schemes known from wired networks? What is done in wired network to avoid this effect?

(R.G.P.V., Dec. 2010)

Ans. Medium access control schemes control user access to the medium using FDM, TDM, CDM, SDM or packet radio access schemes. In wired networks such as Ethernet, the sender senses the medium and waits until it gets the medium free to send the data, and continues to listen to the medium with the help of the carrier sense multiple access with collision detection (CSMA/CD) scheme. Any signal drop can be sensed in a wired network, but this is not possible in the wireless network as the sensing data may collide with another data coming from a hidden terminal.

Hidden and Exposed Terminals – Refer the ans. of Q.9 (i).

Q.13. What is ad-hoc network? Explain the characteristics of it.

(R.G.P.V., Dec. 2006, June 2007, 2011)

Or

Write a detailed note on mobile ad-hoc networks. (R.G.P.V., June 2010)

Ans. Ad-hoc or distributed network topology is used in reconfigurable networks which can operate without the need for a fixed infrastructure. These networks are mainly used by the military and also in a few commercial applications for voice and data transmission. This topology is suitable for rapid deployment of a wireless network in a mobile or fixed environment.

Fig. 3.7 depicts two variations of the ad-hoc network topology. Fig. 3.7(a) is a single-hop ad-hoc network where every user terminal has the functional capability of communicating directly with any of the other user terminals.

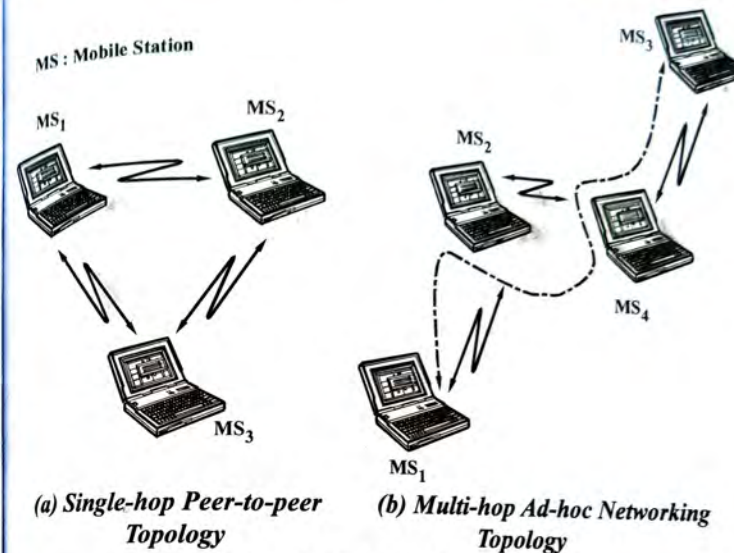


Fig. 3.7 Ad-hoc Networking

In some ad-hoc networking applications, users may be distributed over a wide area and a given user terminal may be able to reach only a portion of the other users in the network owing to transmitter signal power limitations. In such a case, user terminals will have to cooperate in carrying messages across the network between widely separated stations. Networks designed to function this way are known as multi-hop ad-hoc networks and are shown in fig. 3.7 (b). In an ad-hoc multi-hop network, each terminal should be aware of the neighbouring terminals in its coverage range. The multi-hop network configuration was originally applied in military tactical networks where providing reliable communications under unpredictable propagation conditions and over widely varying geographic areas was important.

Characteristics of Ad-hoc Networks – The characteristics of ad-hoc networks are as follow –

- (i) Lack of pre-configuration, meaning network configuration and management must be automatic and dynamic.
- (ii) Node mobility, resulting in constantly changing network topologies.
- (iii) Multi-hop routing.

- (iv) Resource limited devices e.g., laptops, PDAs and mobile phones have power and CPU processing constraints.
- (v) Resource limited wireless communications, e.g., reduced to 10's of kilobits per second by the fact that many nodes must share the radio medium.
- (vi) Potentially large networks, e.g., a network of sensors may comprise thousands or even tens of thousands of mobile nodes.

Q.14. What is Ad-hoc network? Write characteristics and applications of Ad-hoc networks. (R.G.P.V., Dec. 2017)

Ans. Ad-hoc Network and Characteristics – Refer the ans. of Q.13.

Applications – Different ad-hoc communication scenarios will show up, driven by the demand for spontaneous communication. Many different communication layers will support ad-hoc communication layer two for the spontaneous creation of links, layer three for efficient routing and the application layer for service discovery, automatic configuration, authentication etc. Multi-hop ad-hoc technologies can furthermore extend the range of devices or lower interference/extend battery lifetime by transmitting via neighbour nodes instead of transmitting directly to a base station with high transmission power.

Q.15. What are the differences between cellular and ad-hoc networks? (R.G.P.V., June 2007)

Ans. To reduce power consumption and reuse the limited radio spectrum resources, a cellular network was formed. Cell size is one of the factors in the channel reuse rate. Basically, the channel reuse rate in a smaller cell size is higher than the channel reuse rate in a bigger cell size. A cellular architecture would then present a challenge to the frequent handover procedures for a smaller cell size would usually induces a higher handoff frequency.

In addition to cellular networks, the ad-hoc network is another network architecture for wireless networks. The ad-hoc network is a non-infrastructure architecture, in which nodes can access services from one another regardless where they are. The main difference between a cellular environment and ad-hoc network is that the ad-hoc method has no fixed infrastructure, allowing nodes to communicate with one another at any time and anywhere.

Q.16. Compare the ad-hoc topology and infrastructure network topologies. (R.G.P.V., Dec. 2009)

Ans. A number of attributes can be used to compare infrastructure and ad-hoc network topologies.

(i) **Scalability** – In peer-to-peer single hop networks, expansion is always limited to the coverage of the radio transmitter and receiver, and there is no simple way to scale up the network coverage or capacity (wireless traffic) that can be supported by the network. In multi-hop ad-hoc networks, as the number of terminals increases the potential coverage of the network is increased.

However, the traffic handling capacity of the network remains the same. To connect an ad-hoc network to the backbone wired network, one needs to use a proxy server with a wireless connection as a member of the ad-hoc network. In practice all terminals supporting ad-hoc networking operate in a dual mode that also supports infrastructure operation. Wireless infrastructure networks are inherently scalable. To scale up a wireless infrastructure network, the number of BSs or APs is increased to expand the coverage area or to increase the capacity while using the same available spectrum. Therefore, for wide area coverage and for applications with variable traffic loads, infrastructure networks are always used.

(ii) **Flexibility** – Operation of infrastructure networks requires deployment of a network infrastructure which is very often time-consuming and expensive. Ad-hoc networks are inherently flexible and can be set-up instantly. Therefore, ad-hoc networks are always used for temporary applications where flexibility is of prime importance.

(iii) **Controllability** – To coordinate proper operation of a radio network, we need to centrally control certain features such as time synchronization, transmitted power of the mobile stations operating in a certain area, and so on. In an infrastructure network, all these features are naturally implemented in the BS or AP. In an ad-hoc network, implementation of these features requires more complicated structures demanding changes in all terminals.

(iv) **Routing Complexity** – In multi-hop peer-to-peer networks, each terminal should be able to route messages to other terminals. This capability requires each terminal to monitor the existence of other terminals and be able to connect to those available in the immediate neighbourhood. For this, there is a need for a routing algorithm that directs information to the next appropriate terminals. Implementation of these features adds to the complexity of the terminal and the network operation. In infrastructure and peer-to-peer single-hop ad-hoc networks, this problem does not exist.

(v) **Coverage** – In WLANs, coverage of the network is an issue of concern because it has an effect on the selection of the topology. In peer-to-peer single-hop network topology, the maximum distance between two terminals is the range of coverage of the wireless interface used in the terminal. In an infrastructure network, two wireless terminals communicate through an AP or a BS. The maximum distance between two terminals is thus twice the range of coverage of a single wireless modem because the communicating terminals may be located at the edge of the coverage area of the BS or AP. In practice often APs or BSs are fixed in opportunistic locations using elevated mountings that increase the coverage of the wireless modem. This usually results in a maximum coverage distance between two terminals that is greater than twice the coverage distance of the same modem in an ad-hoc configuration.

(vi) **Reliability** – Another issue of concern in small scale WLAN operation and military applications in battlefields is resistance to failure. Infrastructure networks are “single failure point” networks. If the AP or BS fails, the entire communications network is destroyed. This problem does not exist in ad-hoc peer-to-peer configurations.

(vii) **Store and Forward Delay and Media Usage Efficiency** – In peer-to-peer single-hop networks, information is transmitted only once, and there is no store and forward procedure. In the infrastructure topology, we have transmission of data twice, once from the source to the BS/AP and once from the BS/AP to the destination. The BS/AP also should store the message and forward it later. This adds to the delay encountered by the data packets. Multi-hop ad-hoc networks may have several transmissions and several store and forward delays that depend on the instantaneous topology and number of hops required to send the data from the source to the destination.

Q.17. What are the fundamental differences between wired network and ad-hoc wireless networks related to routing? Also discuss difficulties arise in comparison to wired network.

Ans. There are some fundamental differences between the wired and ad-hoc wireless routing.

(i) The main problem for ad-hoc networks is routing arises from the highly dynamic topology. In ad-hoc networks, routing tables must somehow reflect these frequent changes in topology, and routing algorithms have to be adapted. Routing algorithms used in wired networks would either react much too slowly or generate too many updates to reflect all changes in topology. Routing table updates in fixed networks, for example, take place every 30 seconds. This updating frequency might be too low to be useful for ad-hoc networks. Some algorithms rely on a complete picture of the whole network. While this works in wired networks where changes are rare, it fails completely in ad-hoc networks.

(ii) Wired networks, too, have redundant links to survive link failures. However, there is only some redundancy in wired networks, which, additionally, are controlled by a network administrator. In ad-hoc networks nobody controls redundancy, so there might be many redundant links upto the extreme of a completely meshed topology. Routing algorithms for wired networks can handle some redundancy, but a high redundancy can cause a large computational overhead for routing table updates.

(iii) In wired networks links exist only where a wire exists, and connections are planned by network administrators. This is not the case for wireless ad-hoc networks. Links come and go depending on transmission characteristics, one transmission might interfere with another, and nodes might overhear the transmissions of other nodes. Interference creates new problems

by ‘unplanned’ links between nodes – if two close-by nodes forward two transmissions, they might interfere and destroy each other. On the other hand, interference might also help routing. A node can learn the topology with the help of packets it has overheard.

(iv) Routing in wireless ad-hoc networks cannot rely on layer three knowledge alone. Information from lower layers concerning connectivity or interference can help routing algorithms to find a good path.

The following difficulties arise in ad-hoc moves nodes –

(i) Centralized approaches will not really work, because it takes too long to collect the current status and disseminate it again. Within this time the topology has already changed.

(ii) Many nodes need routing capabilities. While there might be some without, at least one router has to be within the range of each node. Algorithms have to consider the limited battery power of these nodes.

(iii) The notion of a connection with certain characteristics cannot work properly. Ad-hoc networks will be connectionless, because it is not possible to maintain a connection in a fast changing environment and to forward data following this connection. Nodes have to make local decisions for forwarding and send packets roughly toward the final destination.

(iv) A last alternative to forward a packet across an unknown topology is flooding. This approach always works if the load is low, but it is very inefficient. A hop counter is needed in each packet to avoid looping, and the diameter of the ad-hoc network, i.e., the maximum number of hops, should be known. (the number of nodes can be used as an upper bound).

Q.18. Name the main differences between multi-hop ad-hoc networks and other networks. What advantages do these ad-hoc networks offer?

(R.G.P.V., Dec. 2010)

Ans. Differences between Multi-hop Ad-hoc Networks and other Networks – Refer the ans. of Q.17.

Advantages of Multi-hop Ad-hoc Networks –

(i) **Disaster Relief** – Infrastructures break down in disaster areas. Hurricanes cut phone and power lines, fires burn servers, floods destroy base stations. Emergency teams can only depend on an infrastructure they can set-up themselves. No forward planning is done, and the set-up must be very fast and reliable. The same applies to many military activities, which is, to be honest, one of the major driving forces behind mobile ad-hoc networking research.

(ii) **Instant Infrastructure** – Spontaneous interpersonal communications, unplanned meetings etc. do not depend on any infrastructure. Infrastructures require planning and administration. It would take too long to set-up this type of infrastructure. Hence, ad-hoc connectivity has to be set-up.

(iii) **Remote Areas** – It is sometimes expensive to set-up an infrastructure in sparsely populated areas even if infrastructures could be planned ahead. Ad-hoc networks or satellite infrastructures can be a solution depending on the communication pattern.

(iv) **Effectiveness** – For certain applications, services provided by existing infrastructures might be too expensive. For example, if only connection-oriented cellular networks exist, but an application sends only a small status information every other minute, a cheaper ad-hoc packet-oriented network might offer a better solution. Registration procedures might take too long, and communication overheads might be too high with existing networks. Application-tailored ad-hoc networks can provide a better solution.

IEEE 802.11 – SYSTEM ARCH, PROTOCOL ARCH, PHYSICAL LAYER, CONCEPT OF SPREAD SPECTRUM, MAC AND ITS MANAGEMENT, POWER MANAGEMENT, SECURITY

Q.19. Explain the overview of the IEEE 802.11. (R.G.P.V., June 2010)
Or

What is IEEE 802.11 ?

(R.G.P.V., Dec. 2004, 2016)

Ans. The IEEE standard 802.11 specifies the most famous family of WLANs in which many products are available. As the standard's number indicates, this standard belongs to the group of 802.x LAN standards. The standard specifies the physical and medium access layer adapted to the special requirements of wireless LANs, but offers the same interface as the others to higher layers to maintain interoperability. The primary goal of the standard was the specification of a simple and robust WLAN which offers time-bounded and asynchronous services.

Additional features of the WLAN should include the support of power management to save battery power, the handling of hidden nodes, and the ability to operate worldwide. The 2.4 GHz ISM band, which is available in most countries around the world was chosen for the original standard. Data rates envisaged for the standard were 1 Mbit/s mandatory and 2 Mbit/s optional.

Q.20. With neat sketch, explain architecture of 802.11 LAN. Also explain its MAC logic.

(R.G.P.V., June 2004, 2008, 2009, 2011)

Ans. Fig. 3.8 shows the model developed by the 802.11 working group. In a wireless LAN, the smallest building block is a basic service set (BSS), which has some number of stations executing the same MAC protocol and competing for access to the same shared wireless medium. A BSS can be isolated or it may connect to a backbone distribution system (DS) through an access point (AP). The access point acts like a bridge. MAC protocol may be

fully distributed or controlled by a control coordination function kept in the access point. Generally, the BSS corresponds to what is referred to as a cell in the literature. The DS can be a switch, a wired network, or a wireless network.

Fig. 3.8 shows the simplest configuration, in which each station belongs to a single BSS. It means that each station is within wireless range only of other stations within the same BSS. It is also possible for two BSSs to overlap geographically, so that a single station could participate in more than one BSS. In addition, the association between a station and a BSS is dynamic. Stations may turn off, come within range, and go out of range.

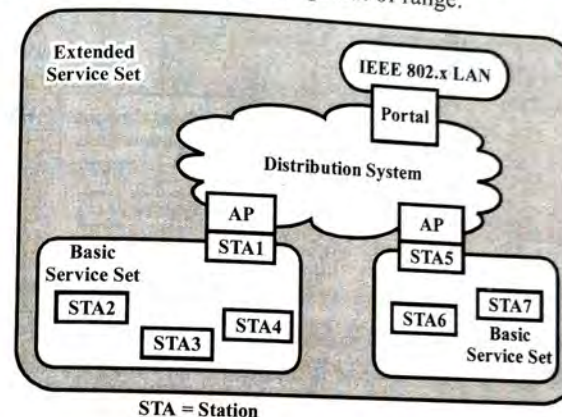


Fig. 3.8 Architecture of 802.11

The extended service set (ESS) has two or more basic service sets interconnected by a distribution system. Typically, the distribution system is a wired backbone LAN but can be any communications network. The extended service set looks like a single logical LAN to the logical link control (LLC) level.

In fig. 3.8, an access point (AP) is implemented as part of a station. The AP is the logic within a station which provides access to the DS by providing DS services in addition to acting as a station. A portal is used to integrate the IEEE 802.11 architecture with a traditional wired LAN. Portal logic is implemented in a device, such as a router or bridge, that is part of the wired LAN and that is attached to the DS.

MAC Logic – IEEE 802.11 medium access control (MAC) logic is shown in fig. 3.9. The distributed coordination function (DCF) sublayer makes use of a simple CSMA (carrier sense multiple access) algorithm. When a station has a MAC frame to transmit, it listens to the medium. In case the medium is idle, the station may transmit, otherwise, the station must wait until the current transmission is complete before transmitting. The DCF does not include a collision detection function (i.e., CSMA/CD) since collision detection is not

practical on a wireless network. Dynamic range of the signals on the medium is too large, so that a transmitting station cannot effectively distinguish incoming weak signals from noise and the effects of its own transmission.

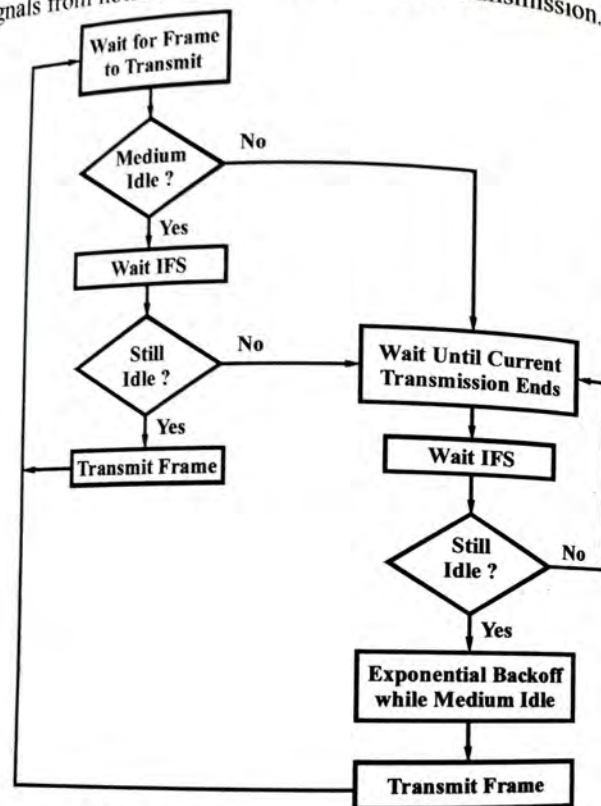


Fig. 3.9 IEEE 802.11 Medium Access Control Logic

The DCF includes a set of delays, that amounts to a priority scheme, to ensure the smooth and fair functioning of this algorithm. We start by considering a single delay known as an interference space (IFS). Actually, there are three different IFS values, but the algorithm is best explained by ignoring this detail.

(i) The medium is sensed by a station which has a frame to transmit. In case, the medium is idle, it waits to see if the medium remains idle for a time equal to IFS. If so, the station may transmit at once.

(ii) In case the medium is busy (either because the station initially finds the medium busy or because the medium becomes busy during the IFS idle time), the station defers transmission and continues to monitor the medium until the current transmission is over.

(iii) After once the current transmission is over, the station delays another IFS. In case the medium remains idle for this period, the station backs off a random amount of time and again senses the medium. If the medium is still

idle, the station may transmit. If the medium is busy during the backoff time, then the backoff timer is halted and resumes when the medium becomes idle.

A technique known as binary exponential backoff is used to ensure that backoff maintains stability. A station will attempt to transmit repeatedly in the face of repeated collisions, but after each collision, the mean value of the random delay is doubled. Binary exponential backoff provides a means of handling a heavy load. Repeated, failed attempt to transmit result in longer and longer backoff times, that helps to smooth out the load. Without such a backoff, following situation could take place. Two or more stations attempt to transmit at the same time, causing a collision. Then, these stations immediately attempt to retransmit, causing a new collision.

Q.21. Explain the architecture of IEEE 802.11. (R.G.P.V., Dec. 2014)

Ans. Refer the ans. of Q.20.

Q.22. Describe 802.11 architecture. How can you achieve wireless LAN security? (R.G.P.V., Dec. 2012)

Ans. 802.11 Architecture – Refer the ans. of Q.20.

Wireless LAN Security – To address the WLAN security issues, a set of capabilities has been developed by the 802.11i task group. To achieve strong security into WLANs, the Wi-Fi Alliance promulgated Wi-Fi Protected Access (WPA) as a Wi-Fi standard. WPA is a set of security mechanisms that removes the security issues of 802.11 and was based on the current state of the 802.11i standard.

Three main security areas are addressed by IEEE 802.11i, namely, authentication, key management, and data transfer privacy. 802.11i needs the use of an authentication server (AS) and defines a more robust authentication protocol to improve authentication. The AS also plays a role in key distribution. For privacy, 802.11i offers three different encryption schemes. The advanced encryption standard (AES) with 128-bit keys is used by the scheme that provides a long term solution.

Q.23. Explain IEEE 802.11a in brief.

Ans. IEEE 802.11a offers upto 54 Mbit/s using OFDM (IEEE 1999), initially aimed at the US 5 GHz U-NII (Unlicensed national information infrastructure) bands. The FCC (US) regulations offer three different 100 MHz domains for the use of 802.11a, each with a different legal maximum power output – 5.15-5.25 GHz/50 mW, 5.25-5.35 GHz/250 mW, and 5.725-5.825 GHz/1 W. ETSI (Europe) defines different frequency bands for Europe – 5.15-5.35 GHz and 5.47-5.725 GHz and requires two additional mechanisms for operation – dynamic frequency selection (DFS) and transmit power control (TPC). Maximum transmit power is 200 mW EIRP for the lower frequency band and 1 W EIRP for the higher frequency band. DFS and TPC are not

necessary, if the transmit power stays below 50 mW EIRP and only 5.15-5.85 GHz are used.

To be able to offer data rates upto 54 Mbit/s IEEE 802.11a uses different technologies. The system uses 52 subcarriers (48 data + 4 pilot) that are modulated using BPSK, QPSK, 16-QAM, or 64-QAM. To mitigate transmission errors, FEC is applied using coding rates of 1/2, 2/3 or 3/4. Table 3.1 gives an overview of the standardized combinations of modulation and coding schemes together with the resulting data rates.

Table 3.1 Rate Dependent Parameters for IEEE 802.11a

Data Rate [Mbit/s]	Modulation	Coding Rate	Coded Bits per Subcarrier	Coded Bits per OFDM Symbol	Data Bits per OFDM Symbol
6	BPSK	1/2	1	48	24
9	BPSK	3/4	1	48	36
12	QPSK	1/2	2	96	48
18	QPSK	3/4	2	96	72
24	16-QAM	1/2	4	192	96
36	16-QAM	3/4	4	192	144
48	64-QAM	2/3	6	288	192
54	64-QAM	3/4	6	288	216

Fig. 3.10 shows the channel layout for the US U-NII bands. The center frequency of a channel is $5000 + 5 \times \text{channel number}$ (MHz). This definition provides a unique numbering of channels with 5 MHz spacing starting from 5.15 GHz. Depending on national regulations, different sets of channels may be used. Eight channels have been defined for the lower two bands in the U-NII (36, 40, 44, 48, 52, 56, 60 and 64); four more are available in the high band (149, 153, 157 and 161). Using these channels allows for interference-free operation of overlapping 802.11a cells. Channel spacing is 20 MHz, the occupied bandwidth of 802.11a is 16.6 MHz.

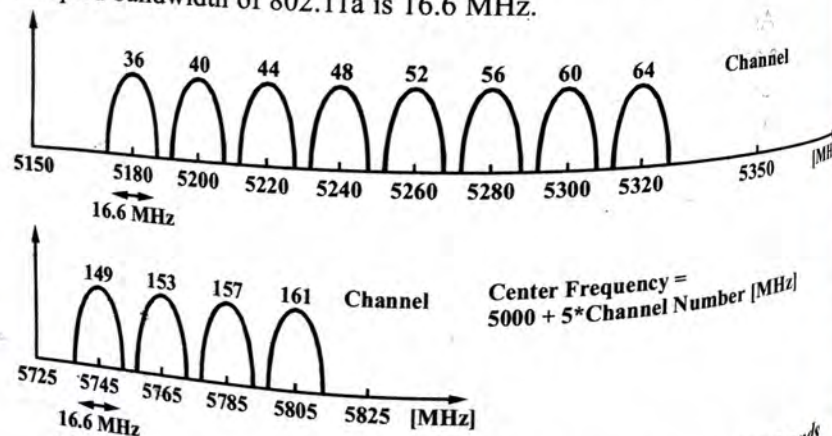


Fig. 3.10 Operating Channels of IEEE 802.11a in the U-NII Bands

Q.24. Explain the basic structure of an IEEE 802.11a.

Ans. The basic structure of an IEEE 802.11a PPDU is shown in fig. 3.11.

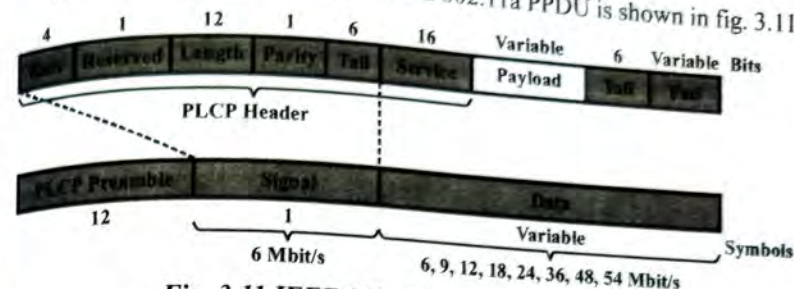


Fig. 3.11 IEEE 802.11a Physical Layer PDU

The **PLCP preamble** consists of 12 symbols and is used for frequency acquisition, channel estimation, and synchronization. The duration of the preamble is 16 μ s.

The following OFDM symbol, called **signal**, contains the following fields and is BPSK-modulated. The 4-bit **rate** field determines the data rate and the modulation of the rest of the packet. The **length** field indicates the number of bytes in the payload field. The **parity** bit will be an even parity for the first 16-bits of the signal field (rate, length and the reserved bit). Finally, the six tail bits are set to zero.

The **data** field is transmit with the rate determined in the rate field and contains a **service** field which is used to synchronize the descrambler of the receiver and which contains bits for future use. The **payload** contains the MAC PDU (1-4095 bytes). The **tail** bits are used to reset the encoder. Finally, the **pad** field ensures that the number of bits in the PDU maps to an integer number of OFDM symbols.

Q.25. Explain IEEE 802.11b in brief.

Ans. As the first commercial 802.11 products came on the market some companies offered proprietary solutions with 11 Mbit/s. A common standard, to avoid market segmentation, IEEE 802.11b soon followed and was added as supplement to the original standard (Higher-speed physical layer extension in the 2.4 GHz band). This standard describes a new PHY layer and is by far the most successful version of IEEE 802.11 available today.

Depending on the current interference and the distance between transmitter and receiver 802.11b systems offer 11, 5.5, 2, or 1 Mbit/s. Maximum user data rate is approximately 6 Mbit/s. The lower data rates 1 and 2 Mbit/s use the 11-chip Barker sequence and DBPSK or DQPSK, respectively. The new data rates 5.5 and 11 Mbit/s use 8-chip **complementary code keying** (CCK).

The standard defines several packet formats for the physical layer. The mandatory format interoperates with the original versions of 802.11. The

optional versions provide a more efficient data transfer due to shorter headers/different coding schemes and can coexist with other 802.11 versions. However, the standard states that control all frames shall be transmitted at one of the basic rates, so they will be understood by all stations in a BSS.

Q.26. Explain the packet format for IEEE 802.11b.

Ans. Fig. 3.12 shows two packet formats standardized for 802.11b. The mandatory format is called **long PLCP PPDU**, in which the rate encoded in the signal field is in multiples of 100 kbit/s. Thus, 0x0A represents 1 Mbit/s, 0x14 is used for 2 Mbit/s, 0x37 for 5.5 Mbit/s and 0x6E for 11 Mbit/s. Note that the preamble and the header are transmitted at 1 Mbit/s using DBPSK. The optional **short PLCP PPDU** format differs in several ways. The short synchronization field consists of 56 scrambled zeros instead of scrambled ones.

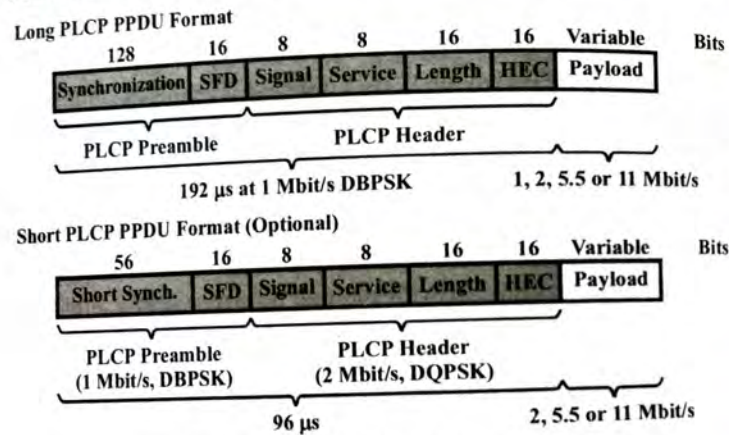


Fig. 3.12 IEEE 802.11b PHY Packet Formats

The short start frame delimiter SFD consists of a mirrored bit pattern compared to the SFD of the long format – 0000 0101 1100 1111 is used for the short PLCP PDU instead of 1111 0011 1010 0000 for the long PLCP PPDU. Receivers that are unable to receive the short format will not detect the start of a frame. Only the preamble is transmitted at 1 Mbit/s, DBPSK. The following header is already transmitted at 2 Mbit/s, DQPSK, which is also the lowest available data rate. The length of the overhead is only half for the short frames, as shown in fig. 3.12. This is useful for e.g. short, but time critical, data transmissions.

Q.27. Write short note on IEEE 802.11g.

Ans. IEEE 802.11g introducing new modulation schemes, forward error correction and OFDM also allows for higher data rates i.e., above 20 Mbit/s at 2.4 GHz. This approach should be backward compatible to 802.11b and should benefit from the better propagation characteristics at 2.4 GHz compared to

5 GHz. Currently, chips for 54 Mbit/s are available as well as first products. An alternative proposal for 802.11g suggests the so called packet binary convolutional coding (PBCC) to reach a data rate of 22 Mbit/s. While the 54 Mbit/s OFDM mode is mandatory, the 22 Mbit/s PBCC mode can be used as an option. The decision between 802.11a and 802.11g is not obvious. Many 802.11a products are already available and the 5 GHz band is (currently) not as crowded as the 2.4 GHz band where not only microwave ovens, but also bluetooth, operate. Coverage is better at 2.4 GHz and fewer access points are needed, lowering the overall system cost. 802.11g access points can also communicate with 802.11b devices as the current 802.11g products show. Dual mode devices will be available covering 802.11a and 802.11 (band g). If a high traffic volume per square metre is expected (e.g., hot spots in airport terminals), the smaller cells of 802.11a access points and the higher number of available channels (to avoid interference) at 5 GHz are clear advantages.

Q.28. Explain various IEEE 802 standards and basic architecture of IEEE 802.11 standard.
(R.G.P.V., Dec. 2011, Nov. 2018)

Ans. Refer the ans. of Q.23, Q.25, Q.27 and Q.20.

Q.29. Draw and explain the architecture of 802.11a, 802.11b and 802.11g.
(R.G.P.V., Dec. 2015)

Ans. Refer the ans. of Q.23, Q.25, Q.26 and Q.27.

Q.30. Draw and explain protocol architecture and bridging for IEEE 802.11.
(R.G.P.V., Dec. 2010)

Ans. An IEEE 802.11 wireless LAN connected to a switched IEEE 802.3 Ethernet through a bridge is shown in fig. 3.13. The WLAN behaves like a

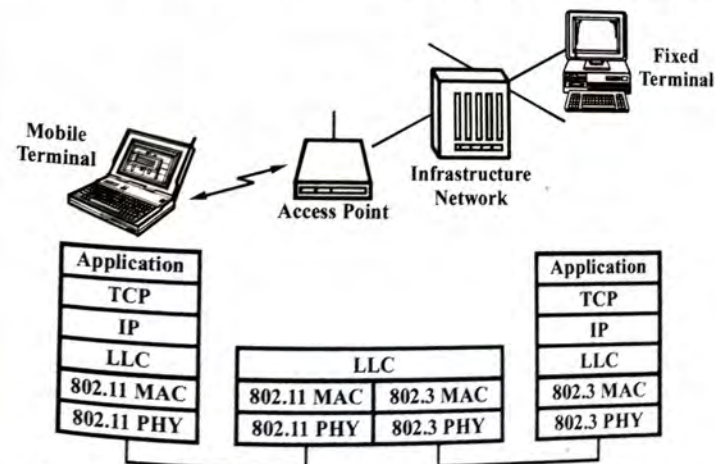


Fig. 3.13 IEEE 802.11 Protocol Architecture and Bridging

slow wired LAN. As a result, the higher layers appear the same for wireless nodes as for wired node. The upper part of the logical link control (LLC), the data link control layer, covers the differences of the medium access control layers required for the different media. No explicit LLC layer is visible in many of today's networks.

The IEEE 802.11 standard covers the medium access layer MAC and the physical layer PHY. The basic tasks of the MAC layer consists of medium access, fragmentation of user data, and encryption. The physical layer is partitioned into the **physical medium dependent** sublayer **PMD** and the **physical layer convergence protocol (PLCP)**. The PMD sublayer controls modulation and encoding/decoding of signals. Finally, the PLCP sublayer provides a carrier sense signal, known as clear channel assessment (CCA), and provides a common PHY service access point (SAP) independent of the transmission technology.

In addition the protocol sublayers, the standard specified management layers and the station management. The MAC management helps the association and reassociation of a station to an access point and roaming between different access points. It also handles encryption, authentication mechanism, power management to save battery power, and synchronization of a station with regard to an access point. It also manages the MAC management information base (MIB).

The PHY management handles channel tuning and PHY MIB maintenance. At last, station management interacts with both management layers. It also provides additional higher layer functions.

Q.31. Explain the IEEE 802.11 protocol architecture with a neat diagram.
(R.G.P.V., Nov. 2019)

Ans. Refer the ans. of Q.30.

Q.32. Explain the functions of various layers of WLAN and the protocol architecture.
(R.G.P.V., Dec. 2011)

Ans. The functions of various layers of WLAN are as follows –

(i) **Physical Layer** – This layer converts a stream of bits into signals that can be transmitted on the sender side. At the receiver, the physical layer transforms the signals back into a bit stream. For wireless communication the physical layer is responsible for frequency selection, generation of the carrier frequency, signal detection, modulation of data onto a carrier frequency and encryption.

(ii) **Data Link Layer** – This layer is responsible for accessing the medium, multiplexing of different data streams, correction of transmission errors, and synchronization. The data link layer is also responsible for a reliable point-to-point connection between two devices or a point-to-multipoint connection between one sender and several receivers.

(iii) **Network Layer** – This layer is responsible for routing packets through a network or establishing a connection between two entities over many other intermediate systems.

(iv) **Transport Layer** – This layer is responsible for establishing an end-to-end connection.

(v) **Application Layer** – Finally, the applications are situated on top of all transmission-oriented layers.

Protocol Architecture – Refer the ans. of Q.30.

Q.33. Describe the physical layer architecture of 802.11 standard.
(R.G.P.V., Dec. 2013)

Ans. Several physical layers are defined by the 802.11 standard to operate with the MAC layer. As shown in fig. 3.14, the physical layer is divided into the physical layer convergence procedure (PLCP) sublayer and the physical medium-dependent (PMD) sublayer.

PCLP is the upper sublayer and communicates with the MAC layer by using the service primitives

through the physical layer service access point (SAP). Based on instructions from the MAC layer, the PCLP

maps a LLC (logical link control) protocol data unit (PDU) into a format which is appropriate for transmission over a given physical

medium. In addition, the LLC PDU also is known as a MAC service data unit (MSDU) because it is the data field at the MAC layer.

A protocol data unit (PDU) consists of a header containing management information, a trailer field, and a block of user information known as a service data unit (SDU).

The PLCP delivers incoming frames from the wireless medium to the MAC layer at a receiving station. As depicted in fig. 3.14, the MAC sublayer adds a header and trailer to the MSDU to form the PLCP SDU (PSDU).

The PLCP delivers incoming frames from the wireless medium to the MAC layer at a receiving station. As depicted in fig. 3.14, the MAC sublayer adds a header and trailer to the MSDU to form the PLCP SDU (PSDU).

The PLCP delivers incoming frames from the wireless medium to the MAC layer at a receiving station. As depicted in fig. 3.14, the MAC sublayer adds a header and trailer to the MSDU to form the PLCP SDU (PSDU).

The PLCP delivers incoming frames from the wireless medium to the MAC layer at a receiving station. As depicted in fig. 3.14, the MAC sublayer adds a header and trailer to the MSDU to form the PLCP SDU (PSDU).

The PLCP delivers incoming frames from the wireless medium to the MAC layer at a receiving station. As depicted in fig. 3.14, the MAC sublayer adds a header and trailer to the MSDU to form the PLCP SDU (PSDU).

The PLCP delivers incoming frames from the wireless medium to the MAC layer at a receiving station. As depicted in fig. 3.14, the MAC sublayer adds a header and trailer to the MSDU to form the PLCP SDU (PSDU).

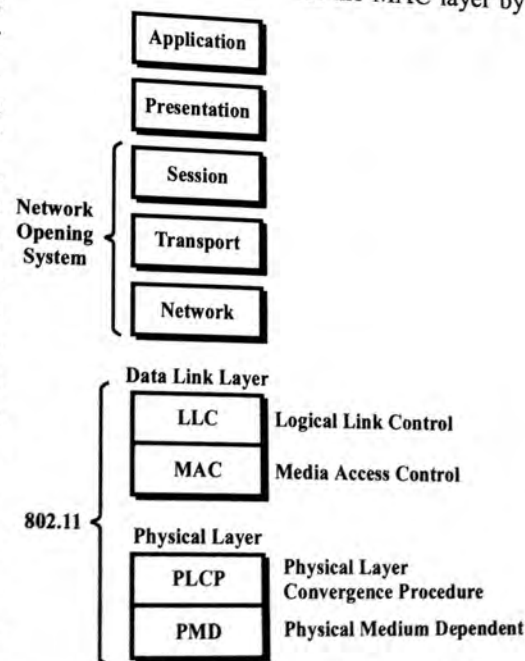


Fig. 3.14 Position of the IEEE-802.11 Standard within the ISO

The PLCP delivers incoming frames from the wireless medium to the MAC layer at a receiving station. As depicted in fig. 3.14, the MAC sublayer adds a header and trailer to the MSDU to form the PLCP SDU (PSDU). The

header has control information and the trailer has a CRC (cyclic redundancy check) field used by the receiver to verify the integrity of the incoming frames. Then, the PLCP sublayer adds two parts to form a PLCP frame or a PLCP PDU (PPDU). First part is a preamble which contains synchronization and start-of-frame information. Second part provides transmission related information. Specific structure of a PLCP frame depends on the particular physical layer with which it is associated.

PMD sublayer is concerned with the characteristics of the physical medium and the specific transmission techniques used to transmit information over a particular medium. In order to carry out the transmission and reception functions, the PLCP and the PMD communicate by means of primitives.

Q.34. Explain the physical layer of IEEE 802.11 standard.

Ans. IEEE 802.11 supports three different physical layers – one layer based on infra red and two layers based on radio transmission. All PHY variants include the provision of the **clear channel assessment** signal (CCA). The CCA is needed for the MAC mechanisms controlling medium access and indicates if the medium is currently idle. The PHY layer offers a service access point (SAP) with 1 or 2 Mbit/s transfer rate to the MAC layer.

The three versions of a PHY layer defined in the standard are described below –

Frequency Hopping Spread Spectrum – FHSS is a spread spectrum technique which allows for the coexistence of multiple networks in the same area by separating different networks using different hopping sequences. The selection of a particular channel for different regions is achieved by using a pseudo-random hopping pattern. National restrictions also determine further parameters, e.g., maximum transmit power is 1 W in the US, 100 mW EIRP in Europe and 10 mW/MHz in Japan.

Fig. 3.15 shows a frame of the physical layer used with FHSS. The frame consists of two basic parts, the PLCP parts (preamble and header) and the payload part. While the PLCP part is always transmitted at 1 Mbit/s, payload, i.e., MAC data, can use 1 or 2 Mbit/s. Additionally, MAC data is scrambled using the polynomial $s(z) = z^7 + z^4 + 1$ for DC blocking and whitening of the spectrum.

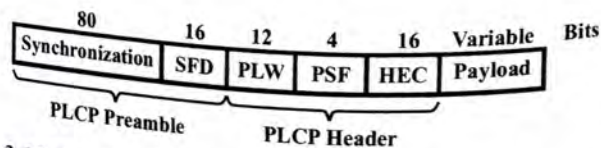


Fig. 3.15 Format of an IEEE 802.11 PHY Frame using FHSS

The frame shown in fig. 3.15 have following fields –

(i) **Synchronization** – The PLCP preamble starts with 80-bit synchronization, which is a 010101 bit pattern. This pattern is used for synchronization of potential receivers and signal detection by the CCA.

(ii) **Start Frame Delimiter (SFD)** – The following 16-bits indicate the start of the frame and provide frame synchronization. The SFD pattern is 0000110010111101.

(iii) **PLCP_PDU Length Word (PLW)** – This first field of the PLCP header indicates the length of the payload in bytes including the 32-bit CRC at the end of the payload. PLW can range between 0 and 4095.

(iv) **PLCP Signalling Field (PSF)** – This 4-bit field indicates the data rate of the payload following. All bits set to zero (0000) indicates the lowest data rate of 1 Mbit/s. The granularity is 500 kbit/s, thus 2 Mbit/s is indicated by 0010 and the maximum is 8.5 Mbit/s (1111). This system obviously does not accommodate today's higher data rates.

(v) **Header Error Check (HEC)** – Finally, the PLCP header is protected by a 16-bit checksum with the standard ITU-T generator polynomial $G(x) = x^{16} + x^{12} + x^5 + 1$.

Direct Sequence Spread Spectrum – DSSS is the alternative spread spectrum method separating by code and not by frequency. In the case of IEEE 802.11 DSSS, spreading is achieved using the 11-chip Barker sequence (+1, -1, +1, +1, -1, +1, +1, +1, -1, -1, -1). The key characteristics of this method are its robustness against interference and its insensitivity to multipath propagation. The implementation of DSSS is more complex compared to FHSS.

IEEE 802.11 DSSS PHY also uses the 2.4 GHz ISM band and offers both 1 and 2 Mbit/s data rates. The system uses differential binary phase shift keying (DBPSK) for 1 Mbit/s transmission and differential quadrature phase shift keying (DQPSK) for 2 Mbit/s as modulation schemes. The symbol rate is 1 MHz, resulting in a chipping rate of 11 MHz. All bits transmitted by the DSSS PHY are scrambled with the polynomial $s(z) = z^7 + z^4 + 1$ for DC blocking and whitening of the spectrum. Many of today's products offering 11 Mbit/s according to 802.11b are still backward compatible to these lower data rates.

Fig. 3.16 shows a frame of the physical layer using DSSS. The frame consists of two basic parts, the PLCP part (preamble and header) and the payload part. While the PLCP part is always transmitted at 1 Mbit/s, payload, i.e., MAC data, can use 1 or 2 Mbit/s.

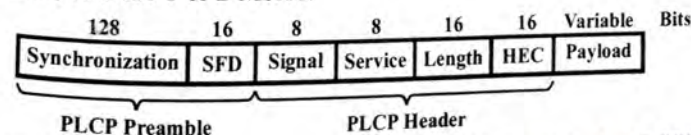


Fig. 3.16 Format of an IEEE 802.11 PHY Frame using DSSS

The frame shown in fig. 3.16 have the following fields –

(i) **Synchronization** – The first 128-bits are not only used for synchronization, but also gain setting, energy detection (for the CCA), and frequency offset compensation. The synchronization field only consists of scrambled 1-bits.

(ii) **SFD** – This 16-bit field is used for synchronization at the beginning of a frame and consists of the pattern 1111001110100000.

(iii) **Signal** – Only two values have been defined for this field to indicate the data rate of the payload. The value 0x0A indicates 1 Mbit/s (DBPSK), 0x14 indicates 2 Mbit/s (DQPSK). Other values have been reserved for future use i.e., higher bit rates.

(iv) **Service** – This field is reserved for future use, however, 0x00 indicates an IEEE 802.11 compliant frame.

(v) **Length** – 16-bits are used in this case for length indication of the payload in microseconds.

(vi) **HEC** – Signal, service, and length fields are protected by this checksum using the ITU-T CRC-16 standard polynomial.

Infrared (IR) – The PHY layer, which is based on IR transmission, uses near visible light at 850-950 nm. IR light is not regulated apart from safety restrictions. The standard does not require a line-of-sight between sender and receiver, but should also work with diffuse light. This allows for point-to-multipoint communication. The maximum range is about 10 m if no sunlight or heat sources interfere with the transmission. Typically, such a network will only work in buildings, e.g., classrooms, meeting rooms etc. Frequency reuse is very simple – a wall is more than enough to shield one IR based IEEE 802.11 network from another. Today, no products are available that offer IR communication based on 802.11. Proprietary products offer, e.g., upto 4 Mbit/s using diffuse infrared light. Alternatively, directed infrared communication based on IrDA can be used.

Q.35. What are main benefits of a spread spectrum system ? Explain direct sequence spread spectrum and frequency hopping spread spectrum. (R.G.P.V., Dec. 2011)

Or

What are the benefits of spread spectrum technique ? Explain DSSS and FHSS techniques. (R.G.P.V., Dec. 2017)

Ans. Spread spectrum techniques involve spreading the bandwidth required to transmit data. There are several benefits of spreading the bandwidth. The main benefit is the resistance to narrowband interference. Fig. 3.17 (a) depicts an idealized narrowband signal from a sender of user data. Now, the

sender spreads the signal in step (b), that is, transforms the narrowband signal into a broadband signal. The energy required to transmit the signal is the same, but it is now spread over a larger frequency range. The power level of the spread signal can be smaller than that of the original narrowband signal without losing data. The power level of the user signal can even be as small as the background noise, depending on the generation and reception of the spread signal.

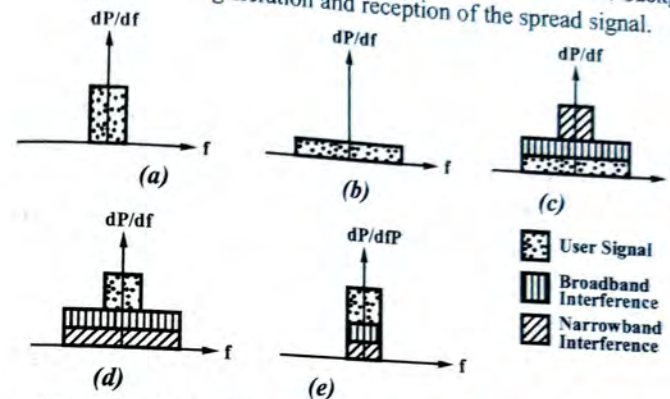


Fig. 3.17 Spreading and Despreading of Spread Spectrum

Narrowband and broadband interference add to the signal in step (c) during transmission. The sum of user signal and interference is received. Now, the receiver knows how to despread the signal, transforming the spread user signal into a narrowband signal again, while spreading the narrowband interference and leaving the broadband interference. In step (e), a bandpass filter is used by the receiver to cut off frequencies left and right of the narrowband signal. At last, the receiver can reconstruct the original data since the power level of the user signal is high.

Spread spectrum is used to deal with narrowband interference for a single channel. It can also be used for many channels. Fig. 3.18 shows a situation. The FDM is used by six different channels for multiplexing. This means that each channel has its own narrow frequency band for transmission. A guard space is required between each frequency to avoid adjacent channel interference. This method needs careful frequency planning. In addition, fig. 3.18 shows a certain channel quality. This is frequency dependent and is a measure for

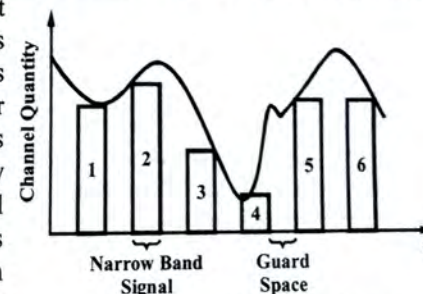


Fig. 3.18 Narrowband Interference without Spread Spectrum

interference at this frequency. Also, channel quality changes over time. Based on receiver characteristics, channels 1, 2, 5 and 6 could be received while the quality of channels 3 and 4 is too bad to reconstruct transmitted data. The transmission of channels 3 and 4 is destroyed by narrowband interference.

Spread spectrum can increase resistance to narrowband interference to help in such a situation. The similar technique is now applied to all narrowband signals. Fig. 3.19 shows that all narrowband signals are now spread into broadband signals using the same frequency range. No more frequency planning is required, and all senders use the same frequency band. But how can receivers recover their signal?

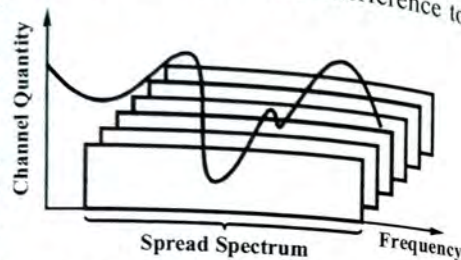


Fig. 3.19 Spread Spectrum to Avoid Narrowband Interference

Now, CDM is used in place of FDM to separate different channels. This application shows the tight coupling of CDM and spread spectrum. Using a special code, spreading of a narrowband signal is achieved. Each channel is allotted its own code, which the receivers apply to recover the signal. The signal cannot be recovered without knowing the code. The signal cannot be recovered, and behaves like background noise. This is the security effect of spread spectrum if a secret code is used for spreading. The coexistence of several signals without coordination, robustness against narrowband interference, relative high security, and a characteristic such as background noise are the features that make spread spectrum and CDM very attractive for military applications.

Direct Sequence Spread Spectrum – Fig. 3.20 shows that direct sequence spread spectrum (DSSS) systems take a user bit stream and perform an XOR with a so called **chipping sequence**. The result is either the sequence 0110101 or its complement 1001010. The chipping sequence composed of smaller pulses, called chips, with a duration t_c , while each user bit has a duration t_b . If the chipping sequence is produced properly it seems as random noise. This sequence is also sometimes known as **pseudo-noise** sequence. The bandwidth of the resulting signal is determined by the spreading factor $s = t_b/t_c$. If the original signal requires a bandwidth w , the resulting signal requires $s \cdot w$ after spreading. Civil applications use spreading factors between 10 and 100, military applications use factors of upto 10,000, while the spreading factor of the very simple example is only 7. For example, wireless LANs complying with the standard IEEE 802.11 use the sequence 10110111000, a so called Barker code, if implemented using DSSS.

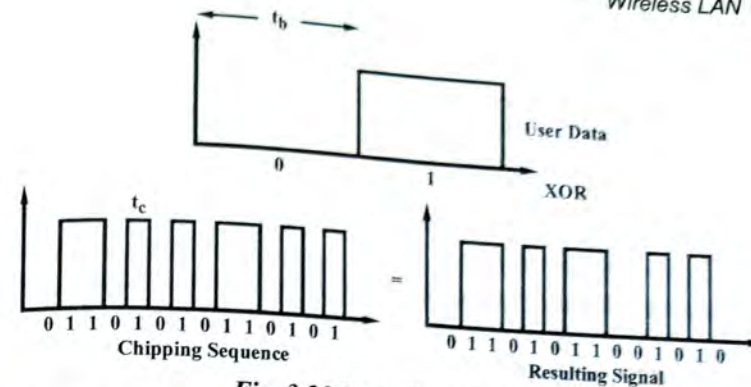


Fig. 3.20 Spreading with DSSS

Fig. 3.21 and fig. 3.22 show the transmitters and receivers using DSSS, respectively. Fig. 3.21 shows that the first step in a DSSS transmitter is the spreading of the user data with the chipping sequence (digital modulation). Then, the spread signal is modulated with a radio carrier (radio modulation). For instance, let a user signal with a bandwidth of 1 MHz. Spreading with the above 11-chip Barker code would result in a signal with 11 MHz bandwidth. This signal is then shifted by the radio carrier to the carrier frequency. Then, this signal is transmitted.

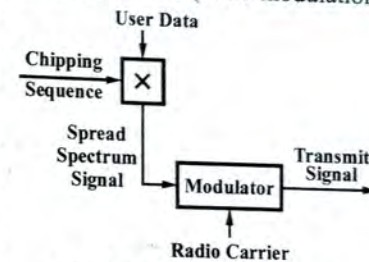


Fig. 3.21 DSSS Transmitter

The DSSS receiver is more complicated compared to the transmitter. The receiver performs the inverse functions of the two transmitter modulation steps. The first step in the receiver is demodulating the received signal. This is done using the same carrier as the transmitter reversing the modulation and results in a signal with approximately the same bandwidth as the original spread spectrum signal. Additional filtering can be used to produce this signal.

The receiver wants to know the original chipping sequence. This means that the receiver produces the same pseudo random sequence as the transmitter. Because the receiver

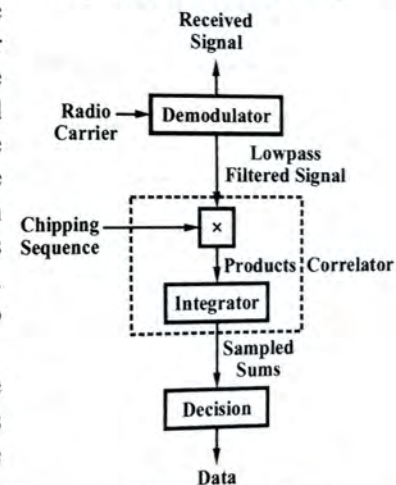


Fig. 3.22 DSSS Receiver

computes the product of a chip with the incoming signal, sequences at the sender and receiver have to be synchronized. This consists of another XOR operation, together with a medium access mechanism that depends on this scheme. An integrator adds all these products during a bit period. Computing the products of chips and signal, and adding the products in an integrator is also known as correlation, the device a correlator. At last, in each bit period a decision unit samples the sums produced by the integrator and determines if this sum represents a binary 0 or 1.

Frequency Hopping Spread Spectrum – The total available bandwidth is divided into several channels of smaller bandwidth plus guard spaces between the channels for **frequency hopping spread spectrum** (FHSS) systems. Receiver and transmitter stay on one of these channels for a certain time and then hop to another channel. This system implements FDM and TDM. The time spent on a channel with a certain frequency is known as the **dwelt time**, the pattern of channel usage is known as the **hopping sequence**.

In slow hopping, one frequency is used by the transmitter for several bit periods. Fig. 3.23 depicts five user bits with a bit period t_b . Performing slow hopping, the frequency f_2 is used by the transmitter for transmitting the first three bits during the dwell time t_d . The transmitter then hops to the next frequency f_3 . Slow hopping systems are cheaper and have relaxed tolerances. But they are not as immune to narrowband interference as fast hopping systems. An option for GSM is slow frequency hopping.

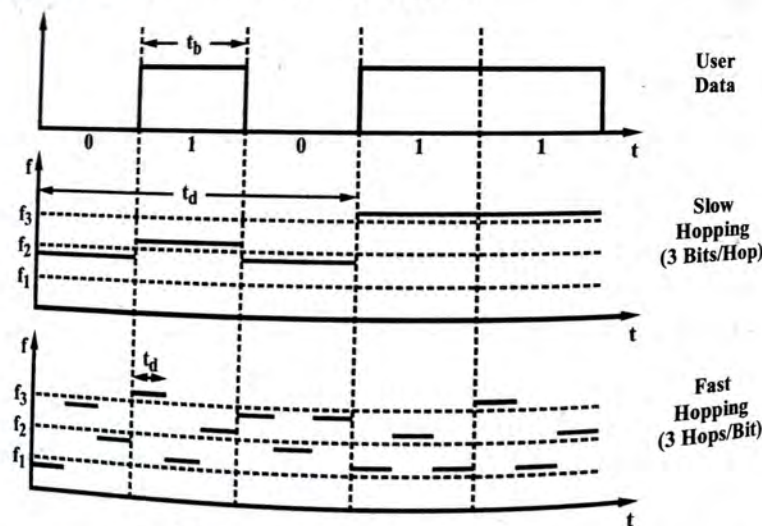


Fig. 3.23 Slow and Fast Frequency Hopping

In fast hopping, the transmitter changes the frequency many times during the transmission of a single bit. In fig 3.23, the transmitter hops three times during a bit period. Fast hopping systems are hard to implement as the

transmitter and receiver have to stay synchronized within smaller tolerances to perform hopping at more or less the same points in time.

The block diagrams of FHSS transmitter and receiver are shown in figs. 3.24 and 3.25 respectively. The modulation of user data using one of the digital-to-analog modulation schemes is the first step. This step provides a narrowband signal, if FSK is used with a frequency f_0 for a binary 0 and f_1 for a binary 1. The next step is to perform frequency hopping on the basis of a hopping sequence. The hopping sequence is fed into a frequency synthesizer producing the carrier frequencies f_i . A second modulation uses the carrier frequency and the modulated narrowband signal to produce a new spread signal with frequency of $f_i + f_1$ for 1 and $f_i + f_0$ for 0 respectively. If distinct FHSS transmissions do not interfere. This needs the coordination of all transmitters and their hopping sequences. Pseudo-random hopping sequences can also be used without coordination. These sequences meet certain properties to keep interference minimal. The same frequency may be chosen by two or more transmitters for a hop, but dwell time is short for fast hopping systems, therefore interference is minimal.

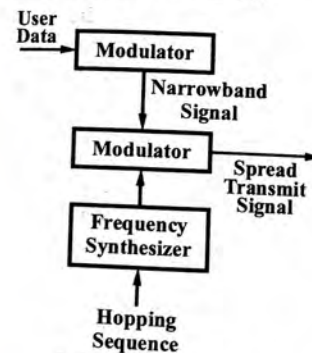


Fig. 3.24 FHSS Transmitter

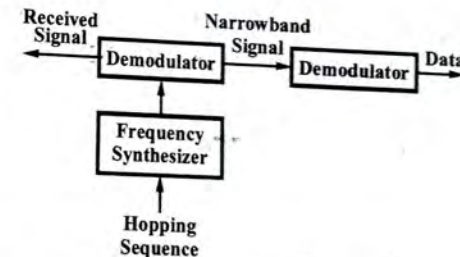


Fig. 3.25 FHSS Receiver

The receiver of an FHSS system must be synchronized and has to know the hopping sequence. Then, it performs the reverse operations of the modulation to reconstruct user data. Also, many filters are required.

Q.36. Write about MAC sublayer.

(R.G.P.V., Dec. 2016)

Ans. The protocols used to determine who goes next on a multiaccess channel belong to a sublayer of the data link layer called MAC (medium access control) sublayer. The MAC sublayer is especially important in LANs, nearly all of which use a multiaccess channel as the basis of their communication. WANs, in contrast, use point-to-point links, except for satellite networks. Because multiaccess channels and LAN are so closely related, including a few issue that are not strictly part of the MAC sublayer.

Technically, the MAC sublayer is the bottom part of the data link layer, so logically we should have studied it before examining all the point-to-point protocols. Nevertheless, for most people, understanding protocols involving multiple parties is easier after two party protocols are well understood. For that reason we have deviated slightly from a strict bottom up order of presentation.

Q.37. With the help of neat diagram, explain the architecture and functions of MAC layer in 802.11 standard.

Ans. MAC sublayer employs one of two methods to gain access to the network. The protocol for these is specified in terms of coordination functions which determine when a station within a BSS may transmit and when it may receive a PDU. First method is distributed coordination function (DCF), that supports asynchronous data transfer on a best-effort basis. By using the DCF, all the stations contend for the channels for each packet transmission. In addition, 802.11 also defines an optional point coordination function (PCF), that uses a centralized decision maker, such as an AP, to regulate transmissions and to provide contention-free frame transfers.

The architecture of these two coordination functions within the MAC sublayer is shown in fig. 3.26. The DCF remains directly on top of the physical layer and the PCF is implemented on top of the DCF. Five physical layer protocols are shown in fig. 3.26. Both the DCF and the PCF can operate concurrently in the same BSS to provide alternating contention and contention-free transmission periods. All stations are needed to support the distributed coordination function. In an ad-hoc network, the stations use only the DCF. In other networks, the stations can operate using just the DCF or a coexisting combination of the DCF and PCF.

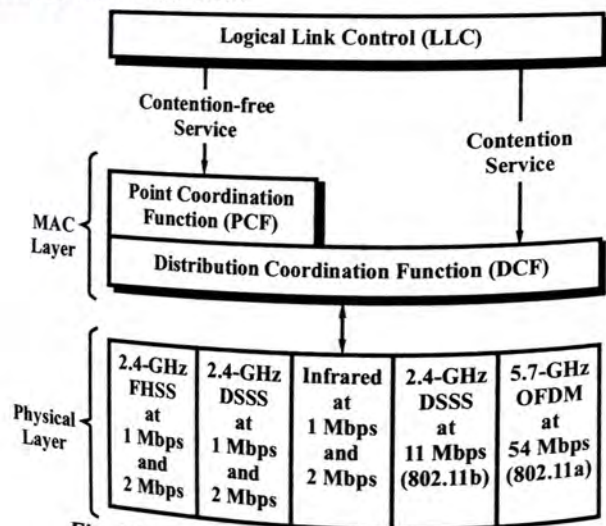


Fig. 3.26 IEEE-802.11 MAC Architecture

Now, we explain these functions in detail.

(i) Distributed Coordination Function (DCF) – DCF is based on a carrier sense multiple access with collision avoidance (CSMA/CA) method, that is similar to the 802.3 Ethernet standard for a wired LAN. But, there is no collision detection function since this is not practical in a wireless network, which is due to the large dynamic range of signals on the medium that makes it difficult to distinguish weak incoming signals from noise. It is obvious that a station must wait for the channel to be idle before transmitting. After once the channel becomes idle, a smooth and fair mechanism must be used to gain access to the channel because other stations also may be waiting to transmit. Otherwise, all waiting stations would transmit at once and collisions would take place, thereby wasting transmission time. To overcome this, the CSMA/CA protocol uses a random set of delays to permit stations to transmit.

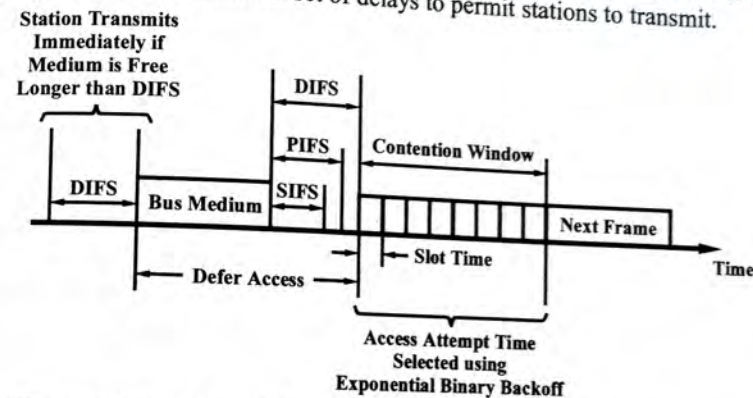


Fig. 3.27 Operation of the Access Method for the CSMA/CA Protocol

Fig. 3.27 shows the operation of the CSMA/CA protocol. After the channel becomes idle, all stations are needed to remain silent for a certain minimum period. This basic time delay is known as the interframe space (IFS). The IFS can take on one of three different length values, depending on the type of frame that the station is about to transmit. Each interval defines the time from the end of the last symbol of the previous frame to the starting of the first symbol of the next frame. Following are the three types of IFS intervals –

(a) Short IFS – Short IFS (SIFS) is the shortest IFS. High-priority frames only require to wait the SIFS period before they contend for the channel. Frames which use the SIFS include acknowledgement (ACK) frames, clear-to-send (CTS) frames, frames which are responding to a poll from the AP, and data frames which are the second or subsequent part of a segmented MAC service data unit.

(b) Point Coordination Function IFS – Point coordination function IFS (PIFS) is intermediate in duration and employed by the PCF to

obtain priority access to the medium. Stations using the PCF have priority over stations using DCF and can send contention-free traffic once they sense that the channel is idle after the PIFS interval.

(c) Distributed Coordination Function IFS – This IFS is the longest IFS. All stations operating under the DCF use the DIFS interval for sending data and management frames. A DCF-based frame has a lower priority than a PCF-based frame because the DIFS is longer than the PIFS.

As shown in fig. 3.27, if the medium is free for a time period longer than the DIFS, the station can transmit immediately. Now, let us consider that the medium is busy and one or more stations operating under the DCF protocol have information to transmit. In this case, the station calculates a random backoff period using a binary exponential backoff scheme to schedule an attempt at gaining access to the channel. This scheduling is done so that all waiting stations have a fair opportunity of transmitting. The time period during which stations attempt to gain access is known as the contention-window. This window is segmented into slots of time.

By using this calculated backoff period, the station sets an internal timer to an integer number of slot times and waits for the medium to become idle. After the end of current transmission, the station waits for a DIFS interval and then senses the medium again. In case the medium is still idle after the DIFS interval, then during the following contention-window interval the station waits for a period equal to the delay time it calculated. If the backoff timer expires and the medium is still idle, then, station may transmit. If the medium becomes busy during the contention-window interval, the backoff procedure is suspended and resumed the next time a contention period occurs.

DCF includes a means for handling lost or corrupted frames. Receiver stations return an ACK (acknowledgement) frame if a data frame is received and its CRC is correct. When the sending station fails to receive such a frame within a specified time period, it will take this as an indication that the frame is corrupted or lost. The time period for each data frame is equal to an ACK frame time plus a SIFS. If an ACK is not received, then the sender executes the backoff procedure to schedule another transmission attempt. If the sender does not receive an ACK frame because the ACK itself was lost, a receiver may get duplicate frames. To detect such duplicates, receiver stations use the sequence numbers.

(ii) Point Coordination Function (PCF) – The priority based PCF can be used to provide contention-free frame transfers. As an example this can be activated to support time-sensitive information such as audio or video. The price for using the PCF is a greater overhead on the network because the PCF requires to send out polling frames that grant a station permission to transmit a frame to any destination. The point coordinator (PC) which performs the

PCF remains in the AP within a BSS. When the PC desires to gain control of the medium, it sends out a beacon at the end of the PIFS interval. The information in the beacon communicates the length of the contention-free period to all stations and prevents them from taking control of the medium until the end of this period.

Q.38. What are the MAC services of IEEE 802.11 that are not provided in the traditional LANs such as 802.3 ?

(R.G.P.V., Dec. 2016)

Ans. The services provided by a IEEE 802.11 network can be divided into two categories – AP services and STA services.

The following are the AP services –

(i) Association – The association service maps the identity of a station to the LAN distribution system via a single AP within a particular BSS. Then, this AP can send the association information to other APs within its affiliated ESS in order to allow the station to roam from one BSS to another.

(ii) Reassociation – The established association is transferred from one AP to another using reassociation. This allow the station to move from one BSS to another.

(iii) Disassociation – A station or an AP may use the diassociation service to terminate an existing association. This is done when nodes leave the BSS or when nodes shut down.

(iv) Distribution – A station applies the distribution service whenever it sends MAC frames across a distribution system. If the destination is in the same BSS, the frame is transmitted directly to the destination, otherwise the frame is sent via the DS.

The following are the STA services –

(i) Authentication – IEEE 802.11 defines an authentication service to control LAN access. All 802.11 stations must use the authentication service to identify themselves prior to establishing a connection with another station. The authentication schemes range from relatively insecure handshaking to public-key encryption schemes.

(ii) Deauthentication – Deauthentication is invoked to terminate existing authentication.

(iii) Privacy – The contents of messages may be encrypted to prevent eavesdroppers from reading the messages.

(iv) Data Delivery – IEEE 802.11 naturally provides a way to transmit and receive data. Although, the transmission is not guaranteed to be completely reliable, like Ethernet.

Q.39. What do you mean by hidden terminal problem? How it is avoided in IEEE 802.11?

Or

Explain MAC-DCF with RTS/CTS extension.

Ans. Hidden terminal problem occurs if one station can receive two others, but those stations cannot receive each other. The two stations may sense the channel is idle, send a frame, and cause a collision at the receiver in the middle. To deal with this problem the standard IEEE 802.11 defines an additional mechanism using two control packets, RTS and CTS. The use of the mechanism is optional, however, every 802.11 node has to implement the functions to react properly upon reception of RTS/CTS control packets.

Fig. 3.28 shows the use of RTS and CTS.

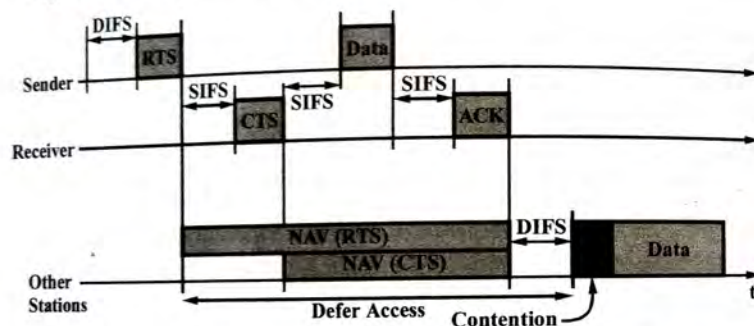


Fig. 3.28 IEEE 802.11 Hidden Node Provisions for Contention-free Access

After waiting for DIFS, the sender can issue a **request to send (RTS)** control packet. The RTS packet thus is not given any higher priority compared to other data packets. The RTS packet includes the receiver of the data transmission to come and the duration of the whole data transmission. This duration specifies the time interval necessary to transmit the whole data frame and the acknowledgement related to it. Every node receiving this RTS now has to set its **net allocation vector (NAV)** in accordance with the duration field. The NAV then specifies the earliest point at which the station can try to access the medium again.

If the receiver of the data transmission receives the RTS, it answers with a **clear to send (CTS)** message after waiting for SIFS. This CTS packet contains the duration field again and all stations receiving this packet from the receiver of the intended data transmission have to adjust their NAV. The latter set of receivers need not be the same as the first set receiving the RTS packet. Now all nodes within receiving distance around sender and receiver are informed that they have to wait more time before accessing the medium.

Finally, the sender can send the data after SIFS. The receiver waits for SIFS after receiving the data packet and then acknowledges whether the transfer was correct. The transmission has now been completed, the NAV in each node marks the medium as free and the standard cycle can start again.

RTS and CTS have been used to avoid the hidden terminal problem, collisions can only occur at the beginning while the RTS is sent. Two or more stations may start sending at the same time. Using RTS/CTS can result in a non-negligible overhead causing a waste of bandwidth and higher delay. An RTS threshold can determine when to use the additional mechanism and when to disable it.

Q.40. Draw the frame format of 802.11 standard and explain it.

Ans. IEEE 802.11 standard defines three classes of frames on the wire – data, control and management. Each of these has a header with a variety of fields used within the MAC sublayer. Also, there are some headers used by the physical layer but these mostly deal with the modulation techniques used.

The format of the data frame is shown in fig. 3.29.

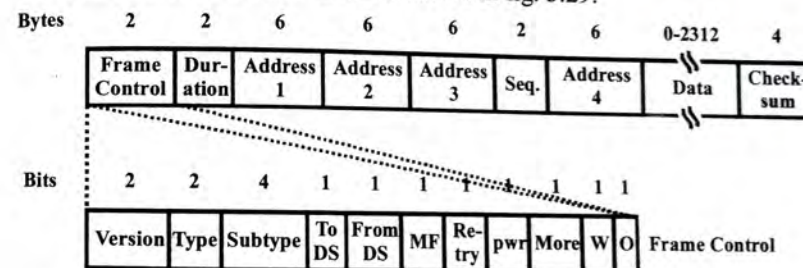


Fig. 3.29 The 802.11 Data Frame

First field is the **frame control** field. It itself has 11 subfields. The first of these is protocol **Version**, which allows to versions of the protocol to operate at the same time in the same cell. Then the **Type** and **Subtype** fields come. The **To DS** and **From DS** bits indicate that the frame is going to or coming from the intercell distribution system. The **MF** bit indicates that more fragments will follow. A retransmission of a frame sent earlier is marked by the **Retry** bit. The **pwr** is stands for power management. It is used by the base station to put the receiver into sleep state or take it out of sleep state.

The **More** bit indicates that the sender has additional frames for the receiver. The **W** bit specifies that the frame body has been encrypted using the **WEP (wired equivalent privacy)** algorithm. Finally, the **O** bit tells the receiver that a sequence of frames with this bit on must be processed strictly in order.

The second field of the data frame, the **Duration** field, shows how long the frame and its acknowledgement will occupy the channel. This field is also present in the control frames and is how other stations manage the NAV

mechanism. The frame header contains four addresses, all in standard IEEE 802 format. Remember that frames may enter or leave a cell via a base station. The other two addresses are used for the source and destination base stations for intercell traffic.

The **Sequence** field allows fragments to be numbered. Of the 16-bits available, 12 identify the frame and 4 identify the fragment. The **Data** field contains the payload, up to 2312 bytes, followed by the usual **Checksum**.

Management frames have a format similar to that of data frames, except without one of the base station addresses, because management frames are restricted to a single cell. Control frames are shorter still, having only one or two addresses, no **Data** field, and no **Sequence** field. The key information here is in the **Subtype** field, usually RTS, CTS, or ACK.

Q.41. Discuss the role of MAC management in IEEE 802.11.

Ans. MAC management plays a central role in an IEEE 802.11 station as it more or less controls all functions related to system integration, i.e., integration of a wireless station into a BSS, formation of an ESS, synchronization of stations etc. The following functions of MAC management are given below –

(i) **Synchronization** – Functions to support finding a wireless LAN, synchronization of internal clocks, generation of beacon signals.

Each node of an 802.11 network maintains an internal clock. To synchronize the clocks of all nodes, IEEE 802.11 specifies a **timing synchronization function (TSF)**. Synchronized clocks are needed for power management, but also for coordination of the PCF and for synchronization of the hopping sequence in an FHSS system. Using PCF, the local timer of a node can predict the start of a super frame, i.e., the contention free and contention period. FHSS physical layers need the same hopping sequences so that all nodes can communicate within a BSS. Within a BSS, timing is conveyed by the periodic transmissions of a beacon frame. A **beacon** contains a timestamp and other management information used for power management and roaming. The timestamp is used by a node to adjust its local clock. The node is not required to hear every beacon to stay synchronized; however, from time to time internal clocks should be adjusted. The transmission of a beacon frame is not always periodic because the beacon frame is also deferred if the medium is busy.

(ii) **Power Management** – Its function is to control transmitter activity for power conservation, e.g., periodic sleep, buffering, without missing a frame.

Wireless devices are battery powered. Therefore, power-saving mechanisms are crucial for the commercial success of such devices. Standard LAN protocols assume that stations are always ready to receive data, although receivers are idle most of the time in lightly loaded networks. However, this

permanent readiness of the receiving module is critical for battery life as the receiver current may be up to 100 mA.

The main idea of IEEE 802.11 power management is to switch off the transceiver whenever it is not needed. For the sending device this is simple to achieve as the transfer is triggered by the device itself. However, since the power management of a receiver cannot know in advance when the transceiver has to be active for a specific packet, it has to 'wake up' the transceiver periodically. Switching off the transceiver should be transparent to existing protocols and should be flexible enough to support different applications. However, throughput can be traded-off for battery life. Longer off-periods save battery life but reduce average throughput and vice versa. The main idea of power saving includes two states for a station – **sleep** and **awake**, and buffering of data in senders. If a sender intends to communicate with a power saving station it has to buffer data if the station is asleep. The sleeping station on the other hand has to wake up periodically and stay awake for a certain time. During this time, all senders can announce the destinations of their buffered data frames. If a station detects that it is a destination of a buffered packet it has to stay awake until the transmission takes place.

(iii) **Roaming** – Functions for joining a network, changing access points, scanning for access points.

(iv) **Management Information Base (MIB)** – All parameters representing the current state of a wireless station and an access point are stored within a MIB for internal and external access. A MIB can be accessed via standardized protocols such as the simple network management protocol (SNMP).

Q.42. Draw and explain the frame format of MAC management.

(R.G.P.V., Dec. 2014)

Ans The MAC management sublayer handles establishment of communications between stations and APs. This layer handles the mechanisms required for a mobile environment. A MAC management frame format is shown in fig. 3.30. A variety of MAC management frame can be used for different purposes.

It starts with frame control field. This field carries instructions on the nature of the packet. It distinguishes data from control and management frame and specifies the type of control or management signalling that the packet is meant to do. The duration/ID field is used to identify the length of the fragmented packets to follow. The destination address (DA) and source address (SA) for the MAC are either two bytes or six bytes but are practically always implemented in six bytes. The sequence control is used for

Frame Control
Duration
DA
SA
BSSID
Sequence Control
Frame Body
FCS

Fig. 3.30 MAC Management Frame Format

fragmentation numbering to control the sequencing. The sequence control and duration/ID are only available in the 802.11 to support fragmentation and the reassembly feature of the MAC protocol. The frame body of the 802.11 contains 0-2312 bytes as compared with 802.3 that have 46-1500 bytes. The frame correction sequence (FCS) field of the DSSS is identical to the CRC field of the FHSS.

Q.43. Explain the MAC management sublayer in term of handoff and power management.
(R.G.P.V., Dec. 2014)

Ans. Handoff – In IEEE 802.11, three kinds of mobility are present, which are given below –

- (i) ESS (extended service set) transition mobility.
- (ii) BSS (basic service set) transition mobility
- (iii) No-transition mobility.

In ESS-transition type of mobility, the mobile station travels from one BSS to another BSS, which is the section of a new ESS. In BSS-transition type of mobility, the mobile station travels from one BSS to another within the same ESS. In no-transition type of mobility, the mobile station is station or travelling within a basic service area.

When a mobile station travels from one BSS to another within the same ESS, the reassociation service is employed. The mobile station always originates it, and it starts the distribution system to identify the fact that the mobile station has travelled its association from one access point (AP) to another. This type of dissociation service is employed to terminate an association.

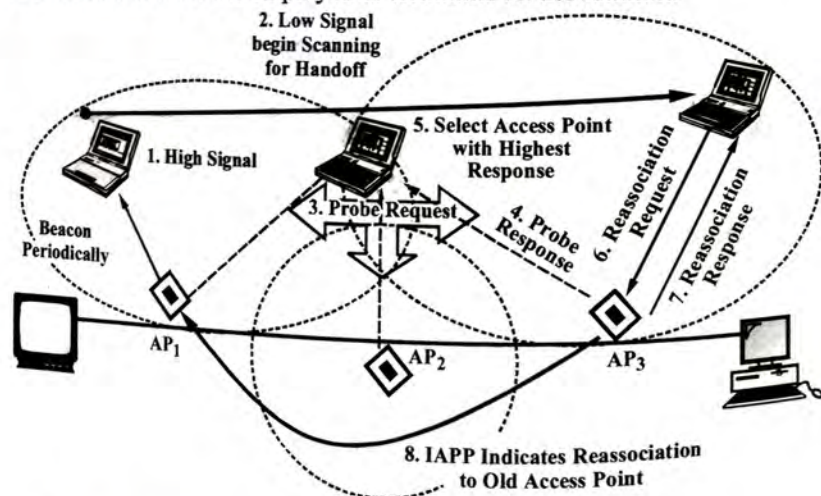


Fig. 3.31

Fig. 3.31 shows the handoff procedures in a Wireless LAN. The access points transmit a beacon signal periodically. The mobile station that powers

on scans the beacon signal and associates itself with the access point with the highest beacon. The beacon signal comprises the information corresponding to the access point like timestamp, capabilities, beacon time interval, ESSID, and traffic indication map (TIM). Various type of access points can be distinguished by the mobile station with the help of information in the beacon.

The mobile station comprises track of the received signal strength (RSS) of the beacon of the access point with which it is associated and mobile station begins to scan from neighbouring APs for stronger beacons, when the RSS signal is low. The scanning process may be either passive or active. The mobile station simply listens to available beacons for the case of passive scanning. The mobile station transmits a probe request to a targeted set of APs which are able to receiving its probe in case of active scanning. Every AP which receives the probe responds with a probe response which comprises the same information which is present in a regular beacon with the exception of the traffic indication map. Hence, the probe response uses as a solicited beacon. The mobile station selects the AP with the highest beacon or probe response. The mobile station transmits a reassociation request to the new AP. The information about the mobile station as well as the old AP are contained in reassociation request. In output stage, the new access point transmits a reassociation response which contains the information about supported station ID, bit rates etc. required to complete resume communication. About the location change, the old access point is not informed by the mobile station. This type of problem can be eliminated with the help of inter access point protocol (IAPP) standard for completing the handoff procedure last stage. The inter access point protocol uses two PDUs to show that a handoff has occurred. As shown in fig. 3.32 these PDUs are used for transferring from the new access point to the old access point over the wired

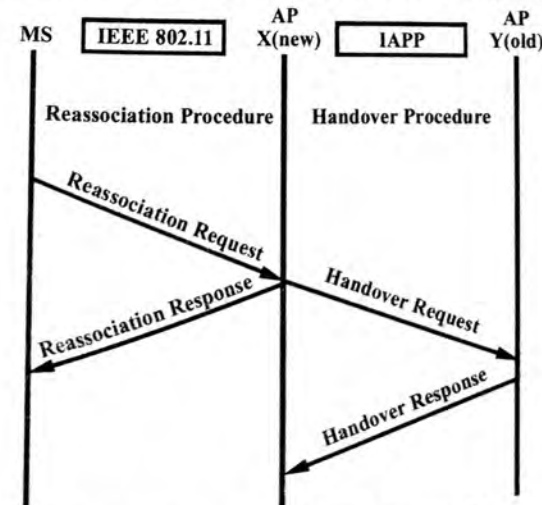


Fig. 3.32 Role of Inter Access Point Protocol in Handoff

network with the help of UDP-IP. An 802.11 subnetwork access protocol (SNAP) is transferred the PDUs, when the access point does not contain an IP address.

Power Management – In wireless LANs, the power conservation difficult is that stations obtain data in bursts but remain in an idle state constantly, which dominates the local area network adaptor power consumption. The problem is how we can power off in the idle periods of time and handle the session. The IEEE 802.11 solution is to introduce the mobile station in sleeping mode, buffer the data at access point and transmit the data when the mobile station is not in a sleep mode.

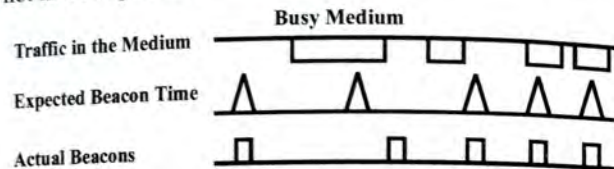


Fig. 3.33

All mobile stations are synchronized with the help of timing synchronization function (TSF), and they will wake up at the same time to listen to the beacon, as shown in fig. 3.33. The power management bit is used by the mobile station in the field of frame control to announce its awake/sleep mode. A traffic indication map (TIM) with every beacon is transmitted that contain the list of stations having buffered data. The mobile station learn that it has a buffered data with the help of checking beacon and traffic indication map (TIM). The mobile station with buffered data transmits a power-save poll to access point. If the mobile station is in active mode, then access point transmits the buffered data.

Q.44. Write short note on security management.

(R.G.P.V., Dec 2005, June 2006, Dec. 2009)

Ans. Security management is both a technical and an administrative consideration in information management. It involves securing access to the network and information flowing in the network, access to data stored in the network, and manipulating the data stored and flowing through the network. The scope of the network and access to it not only covers an enterprise intranet network, but also the Internet, to which it is connected.

Another area of concern in secure communication is the use of mobile stations. You may be familiar with the embarrassing case of politicians conversation from a car phone being interpreted by a third party travelling in an automobile. Although that was an analog signal, the same thing could happen in the case of mobile digital station such as a hand-held stock trading device. An intruder could intercept messages and alter trade transactions either to benefit by them or to hurt the person sending or receiving them.

The goal of security management is to control access to network resources so that the network cannot be sabotaged and sensitive information cannot be accessed without appropriate authorization.

Security management consists of –

- (i) Identifying the sensitive information to be protected.
- (ii) Finding the access points.
- (iii) Securing the access points.
- (iv) Maintaining the secure access points.

Identifying the sensitive information means determining what sensitive information there is on the network and on which hosts it resides. Most services use or offer sensitive information, but there are also services that offer harmless information. When the sensitive information is identified, the next step is to find out how users can access it. This could, for example, be done by port scanning the resources. The result of this is that you know every access point there is to sensitive information. When we know this, the access points can be secured. The security can be applied by –

- (i) Encrypting the traffic.
- (ii) Restricting what traffic can flow on the network, using packet filters.
- (iii) Authentication to services.

Username, passwords, configuration and other sensitive data flowing on the network should be encrypted and if someone eavesdrops on the wire, the encrypted information would be useless.

The network management system itself can contain sensitive information. When determining the access points to a data network, you should also include the network management system.

When the access points are secured they have to be maintained. New security bugs appear constantly and you need to check if your network is vulnerable. There are programs that scan networks for security vulnerabilities and if the network engineer does this every now and then, the risk for intrusion is much smaller. New software patches might be released and installing those as soon as possible is also an important issue when it comes to maintaining the security of a network.

MOBILE IP – UNSUITABILITY OF TRADITIONAL IP, GOALS, TERMINOLOGY, AGENT ADVERTISEMENT AND DISCOVERY, REGISTRATION, TUNNELING TECHNIQUES

Q.45. What is mobile IP? Explain the basic operation of mobile IP.

(R.G.P.V., June 2011, Nov. 2018)

Or

What do you understand by mobile IP? Discuss its basic operation.

(R.G.P.V., Dec. 2018)

Or

Draw a figure and explain the operation of mobile IP.

(R.G.P.V., Dec. 2015)

Ans. Mobile IP – Mobile IP was developed to enable computers to maintain Internet connectivity while moving from one Internet attachment point to another. Mobile IP is particularly suited to wireless connections. However, mobile IP can work with wired connections, in which a computer is unplugged into another.

In this context, the term mobile implies that a user is connected to one or more applications across the Internet, that the user's point of attachment changes dynamically, and that all connections are automatically maintained despite the change. This is in contrast to a user, such as a business traveler, with a portable computer of some sort who arrives at a destination and uses the computer's notebook to dial into an ISP (Internet service provider). In this latter case, the Internet connection of the user is terminated each time the user moves and a new connection is initiated when the user dials back in. Each time an Internet connection is established, software in the point of attachment (typically an ISP) is used to achieve a new, temporarily assigned IP address. This temporary IP address is used by the user's correspondent for each application level connection (for example, FTP, Web connection).

Operation of Mobile IP – In an IP datagram, routers make use of the IP address to perform routing. Particularly, the network portion of an IP address is used by routers to move a datagram from the source computer to the network to which the target computer is attached. Then, the final router on the path, that is attached to the same network as the target computer, uses the host portion of the IP address to deliver the IP datagram to the destination. In addition, this IP address is known to the next higher layer in the protocol architecture. Particularly, most applications over the Internet are supported by TCP connections. If a TCP connection is set-up, then the TCP entity on each side of the connection knows the IP address of the correspondent host. If a TCP segment is handed down to the IP layer for delivery, TCP provides the IP address and IP creates an IP datagram with that IP address in the IP header and sends the datagram out for routing and delivery. However, the IP address may change with a mobile host while one or more TCP connections are active.

Fig. 3.34 illustrates in general terms how mobile IP deals with the problem of dynamic IP addresses. A mobile node is allocated to a particular network, known as its home network. Its IP address on that network, known as its home address, is static. If the mobile node moves its attachment point to another network, that is considered a foreign network for this host. Once the mobile node is reattached, it makes its presence known by registering with a network node, typically a router, on the foreign network called a foreign agent. Then, the mobile node communicates with a similar agent on the user's home network,

called a home agent, giving the home agent the care-of address of the mobile node, the care-of address recognizes the location of the foreign agent. Typically, one or more routers on a network will implement the roles of both home and foreign agents.

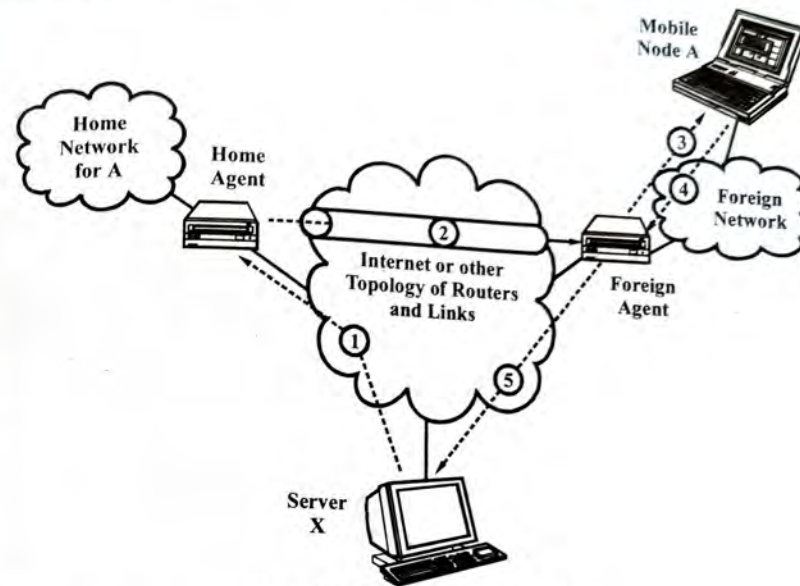


Fig. 3.34 Mobile IP Scenario

When IP datagrams are exchanged over a connection between the mobile node and another host (a server in fig. 3.34), the following operations take place –

- (i) Server X sends an IP datagram destined for mobile node A, with A's home address in the IP header. The IP datagram is routed to A's home network.
- (ii) The incoming IP datagram is intercepted by the home agent at the home network. The home agent encapsulates the whole datagram inside a new IP datagram which has the A's care-of address in the header, and resends the datagram. The use of an outer IP datagram with a different destination IP address is called tunnelling. This IP datagram is routed to the foreign agent.
- (iii) Foreign agent strips off the outer IP header, encapsulates the original IP datagram in a network-level PDU (for example, a LAN LLC frame), and delivers the original datagram to A across the foreign network.
- (iv) If A transmits IP traffic to X, it uses X's IP address. In this example, this is a fixed address, that is, X is not a mobile node. Each IP datagram is transmitted by A to a router on the foreign network for routing to X. Typically, this router is also the foreign agent.
- (v) Using X's IP address, the IP datagram from A to X travels directly across the Internet to X.

Mobile IP includes the following three basic capabilities to support the operations (shown in fig. 3.34) –

- (i) **Discovery** – A mobile node employs a discovery procedure to recognize perspective home agents and foreign agents.
- (ii) **Registration** – Mobile node employs an authenticated registration procedure to inform its home agent of its care-of address.
- (iii) **Tunnelling** – To forward IP datagrams from a home address to a care-of address, tunnelling is used.

Q.46. Explain how mobile Internet (mobile IP) protocol works.

(R.G.P.V., Dec. 2016)

Ans. Refer the ans. of Q.45.

Q.47. What are the requirements associated with the mobile IP standard? How they are met by mobile IP?

(R.G.P.V., June 2010)

Or

Name the requirements for a mobile IP and justify them. Does mobile IP fulfill them all.

(R.G.P.V., Dec. 2011)

Or

What are the requirements of mobile IP?

(R.G.P.V., Dec. 2014)

Ans. Several field trials and proprietary systems led to mobile IP as standard to enable mobility in the Internet. Following requirements accompanied the development of the standard –

(i) **Transparency** – For many higher layer protocols and applications, mobility should remain ‘invisible’. Besides may be noticing a lower bandwidth and some interruption in service, higher layers should continue to work even if the mobile computer has changed its point of attachment to the network. For TCP, it means that the computer must keep its IP address. When the interruption of the connectivity does not take too long, TCP connections survive the change of the attachment point. Obviously, many of today’s applications have not been designed for use in mobile environments, therefore the only effects of mobility should be a higher delay and lower bandwidth. However, there are some applications (for example, cost-based routing or video compression) for which it is better to be ‘mobility aware’. Knowing that it is currently possible to use different networks, the software could select the cheapest one. Additional mechanisms are necessary to inform these applications about mobility.

(ii) **Security** – Mobility poses several security problems. The minimum requirement is that of all the messages related to the management of mobile IP are authenticated. When the IP layer forwards a packet to a mobile host, it must be sure that this host receives the packet. IP layer can only guarantee that the IP address of the receiver is correct. There are no ways of preventing fake IP addresses or other attacks. This is left to higher layers.

(iii) **Compatibility** – The installed base of Internet computers, i.e., computers running TCP/IP and connected to the Internet, is very vast. A new standard cannot introduce changes for applications or network protocols already in use. Mobile IP has to be integrated into existing operating systems or at least work with them. Routers within the Internet should not necessarily need other software. While it is possible to increase the capabilities of some routers to support mobility, it is almost impossible to alter all of them. Mobile IP has to remain compatible with all lower layers used for the standard, non-mobile IP. Mobile IP must not need special media or MAC/LLC protocols, therefore it must use the same interfaces and mechanisms to access the lower layers as IP does. Lastly, end-systems enhanced with a mobile IP implementation should still be able to communicate with fixed systems without mobile IP. Mobile IP has to ensure that users can still access all the other servers and systems in the Internet, but that implies using the same address format and routing mechanisms.

(iv) **Scalability and Efficiency** – Introducing a new mechanism to the Internet must not jeopardize its efficiency. Enhancing IP for mobility must not produce too many new messages flooding the entire network. A great care has to be taken considering the lower bandwidth of wireless links.

Several mobile systems will have a wireless link to an attachment point, therefore only some additional packets should be essential between a mobile system and a node in the network. Looking at the number of computers connected to the Internet and at the growth rates of mobile communication, it is obvious that myriad devices will take part in the Internet as mobile components. It is important for a mobile IP to be scalable over a large number of participants in the entire Internet, worldwide.

Q.48. Explain mobile IP in terms of goals, assumptions and requirements.

(R.G.P.V., Dec. 2010)

Ans. Mobile computing is obviously the paradigm of the future. The network for global data communication with hundreds of millions of users is the Internet. So, why not use a mobile computer in the Internet?

The reason is simple – we will not get a single packet as soon as we leave our home network, that is, the network our computer is configured for, and reconnect our computer at another place. The reason for this is simple if we consider routing mechanisms on the internet. An IP packet is sent by a host with the header containing a destination address with other fields. The destination address determines the receiver of the packet, and the physical subnet of the receiver. For instance, the destination address 129.13.42.99 specifies that the receiver must be connected to the physical subnet with the network prefix 129.13.42. In the internet, routers now look at the destination addresses of incoming packets and forward them according to internal lookup tables. Only prefixes are stored and further optimizations are applied to

avoid an explosion of routing tables. Otherwise, a router would have to store the address of all computers in the internet, which is clearly not feasible. As long as the receiver can be reached within its physical subnet, it gets the packets, as soon as it moves outside the subnet, a packet will not reach it. A host requires a so-called **topologically correct address**.

A quick solution to this problem would be to assign to the computer a new, topologically correct IP address. This is what several users do using DHCP. So moving to a new location would mean assigning a new IP address. The problem is that no one knows about this new address. It is almost impossible to get a host on the Internet, which has just changed its address.

One could say that using dynamic DNS an update of the mapping logical name – IP address is possible. This is what several users do if they have a dynamic IP address.

So what about dynamically adapting the IP address with regard to the current location. The problem is that the DNS requires some time before it updates the internal tables required to map a logical name to an IP address. If the mobile node moves quite often, this approach does not work. The DNS and Internet have not been built for frequent updates. Just suppose millions of nodes moving at the same time. DNS could never provide a consistent view of names and addresses, because it uses caching to improve scalability.

There is a serious problem with higher layer protocols like TCP which rely on IP addresses. Changing the IP address while a TCP connection is still open means breaking the connection. A TCP connection is identified by the tuple (source IP address, source port, destination IP address, destination port) also referred to as a **socket pair**. Hence, a TCP connection cannot tolerate any address change. Breaking TCP connections is not an option, using even simple programs such as telnet would be impossible. The mobile node would also have to notify all communication partners about the new address.

The creation of specific routes to the mobile node is another approach. Routers always select the best-fitting prefix for the routing decision. Now, if a router has an entry for a prefix 129.13.42 and an address 129.13.42.99, it would select the port associated with the latter for forwarding, if a packet with the destination address 129.13.42.99 comes in.

Requirements – Refer the ans. of Q.47.

Q.49. Explain discovery, registration, and tunnelling operations in mobile IP. (R.G.P.V., June 2004)

Ans. Discovery – In mobile IP, the discovery process is very similar to the router advertisement process defined in ICMP. Accordingly, with one or more extensions specific to mobile IP, agent discovery makes use of ICMP router advertisement messages. For an ongoing discovery process, the mobile

node is responsible and it must determine if it is attached to its home network or in which case IP datagrams may be received without forwarding or it is attached to a foreign network. Since handoff from one network to another takes place at the physical layer, a transition from the home network to a foreign network can take place at any time without notification to the network layer (i.e., the IP layer). Hence, for a mobile node, discovery is a continuous process.

For the purpose of discovery, a router or other network node which can act as an agent periodically issues a router advertisement ICMP message with an advertisement extension. The router advertisement portion of the message has the IP address of the router. The advertisement extension includes additional information about the role of the router as an agent. A mobile node hears for these agent advertisement messages. Since a foreign agent could be on the home network of the mobile node (set-up to serve visiting mobile nodes), the arrival of an agent advertisement does not necessarily tell the mobile node that it is on a foreign network. The mobile node must compare the network portion of the router's IP address with the network portion of its own home address. In case these network portions do not match, the mobile node is on a foreign network.

Agent advertisement extension, following the ICMP router advertisement fields, consists of the following fields (refer fig. 3.35) –

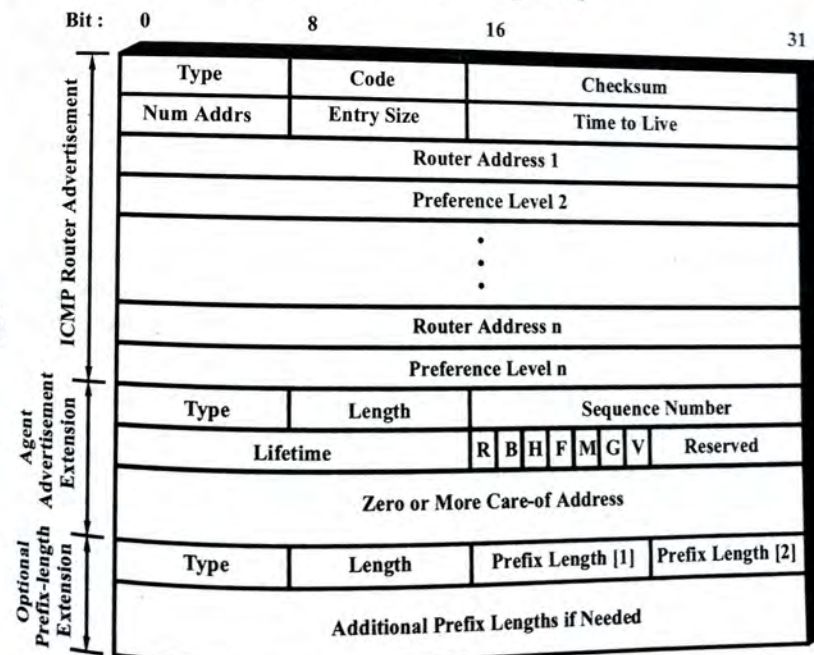


Fig. 3.35 Mobile IP Agent Advertisement Message

- (i) **Type** – 16, shows that this is an agent advertisement.
- (ii) **Length** – $(6 + 4N)$, where N is the number of care-of addresses advertised.
- (iii) **Sequence Number** – Count of agent advertisement messages transmitted because the agent was initiated.
- (iv) **Lifetime** – Longest lifetime, in seconds, that this agent is willing to accept a registration request from a mobile node.
- (v) **R** – Registration with this foreign agent is needed. Even those mobile nodes that have already acquired a care of address from this foreign agent must re-register.
- (vi) **B** – Busy. Foreign agent will not accept registrations from additional mobile nodes.
- (vii) **H** – This agent provides services as a home agent on this network.
- (viii) **F** – This agent provides services as a foreign agent on this network.
- (ix) **M** – This agent can receive tunnelled IP datagrams which use minimal encapsulation.
- (x) **G** – This agent can receive tunnelled IP datagrams which use GRE encapsulation.

(xi) **V** – This agent supports the use of Van Jacobson header compression. This algorithm is defined in RFC 1144 for compressing fields in the TCP and IP headers.

(xii) **Care-of Address** – Care-of Address or addresses supported by this agent on this network. If the **F** bit is set, then there must be at least one such address. There may be multiple addresses.

Registration – Having received a COA, the MN has to register with the HA. The registration is used to inform the HA of the current location for correct forwarding of packets. Depending on the location of the COA, registration can be done in two different ways –

- (i) When the COA is at the FA, the registration is done as shown in fig. 3.36. The MN transmits its registration request containing the COA (refer fig. 3.37) to the FA that is forwarding the request to the HA. Now, the HA sets up a mobility binding containing the home IP address of the mobile node and the current COA. Furthermore, the mobility binding has the lifetime of the registration that is negotiated during the registration process. After the lifetime, registration expires automatically and is deleted, therefore, an MN should register before expiration. This mechanism is essential to avoid mobility bindings that are no longer used. When the mobility binding set-up, the HA transmits a reply message back to the FA that forwards it to the MN.

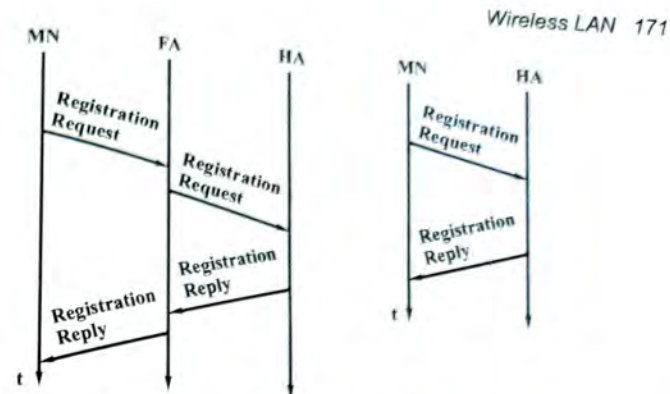


Fig. 3.36 Registration of a Mobile Node via the FA or Directly with the HA

(ii) When COA is co-located, registration can be simpler, as depicted in fig. 3.37. The MN may transmit the request directly to the HA and vice-versa. By the way, this is also the registration procedure for MNs returning to their home network. Here, they also register directly with the HA. But, in case the MN received an agent advertisement from the FA, it should register via this FA if the **R** bit is set in the advertisement.

Advertisement																		
0	7						8	15						16	23		24	31
Type 1							S	B	D	M	G	r	T	x	Lifetime			
Home Address																		
Home Agent																		
COA																		
Identification																		
Extensions.....																		

Fig. 3.37 Registration Request

For registration requests, UDP packets are employed. IP source address of the packet is set to the interface address of the MN, the IP destination address is that of the FA or HA. The UDP destination port is set to 434. In wireless environments, UDP is used due to low overheads and better performance compared to TCP. The fields relevant for mobile IP registration requests follow as UDP data as shown in fig. 3.38. The fields are defined as below –

0	7	8	15	16	31
Type = 3		Code		Lifetime	
Home Address					
Home Agent					
Identification					
Extensions.....					

Fig. 3.38 Registration Reply

For a registration request, the first field type is set to 1. With the S bit, an MN can specify if it desires the HA to retain prior mobility bindings. This permits for simultaneous bindings. The following bits represent the requested behaviour for packet forwarding. Generally, setting the B bit shows that an MN also wants to receive the broadcast packets that have been received by the HA in the home network. When an MN employs a co-located COA, it also takes care of the decapsulation at the tunnel endpoint. The D bit denotes this behaviour. The following bits M and G indicate the use of minimal encapsulation or generic routing encapsulation, respectively. T denotes reverse tunnelling, r and x are set to zero.

Lifetime represents the validity of the registration in seconds. A value of zero shows deregistration. All bits set denotes infinity. The home address is the fixed IP address of the MN, home agent is the IP address of the HA, and COA denotes the tunnel endpoint. To recognize a request and match it with registration replies, the 64-bit identification is produced by the MN. This field is employed for protection against replay attacks of registrations. Extensions must at least have parameters for authentication. A registration reply, that is conveyed in a UDP packet, has a type field set to 3 and a code showing the result of the registration request.

Tunnelling – After once a mobile node is registered with a home agent, the home agent must be able to intercept IP datagrams transmitted to the home address of the mobile node so that these datagrams can be forwarded via tunnelling. For this purpose, the standard does not mandate a specific technique but references ARP (address resolution protocol) as a possible mechanism. The home agent requires to inform other nodes on the same network that IP datagrams with a destination address of the mobile node in question should be delivered to this agent. The home agent steals the identity of the mobile node in order to capture packets destined for that node which are sent across the home network.

As an example, assume that R3 in fig. 3.39 is acting as the home agent for a mobile node which is attached to a foreign network elsewhere on the Internet. It means that there is a host H whose home network is LAN Z which is now attached to some foreign network. When host D contains traffic for H, it will create an IP datagram with H's home address in the IP destination address field. The IP module in D identifies that this destination address is on LAN Z and therefore passes the datagram down to the link layer with instructions to deliver it to a particular MAC-level address on Z. Prior to this time, R3 has informed the IP layer at D that datagrams destined for that particular address should be transmitted to R3. Hence, the MAC address of R3 is inserted by D in the destination MAC address field of the outgoing MAC frame. Likewise, when an IP datagram with the home address of the mobile node arrives at router R2, it identifies that the destination address is on LAN Z and will attempt to deliver the datagram to a MAC-level address on Z. Again, R2 has previously

been informed that the MAC-level address it requires corresponds to R3. For traffic which is routed across the Internet and arrives at R3, for this destination address. R3 must simply identify that the datagram is to be captured and forwarded for this destination address.

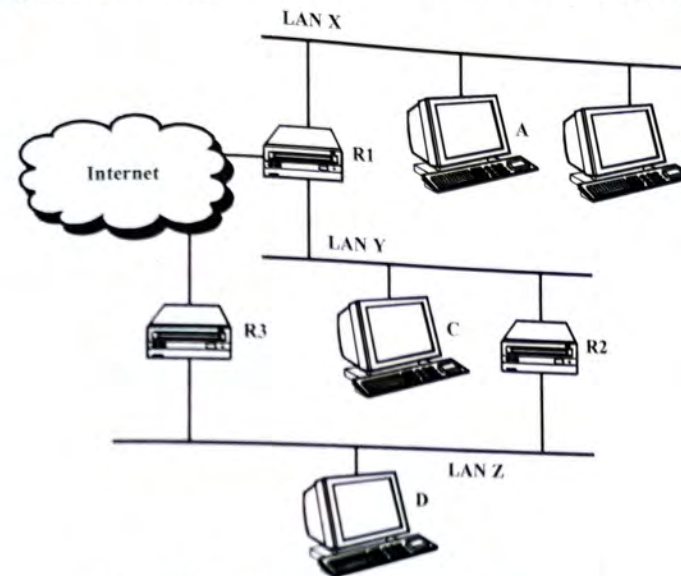


Fig. 3.39 A Simple Example of Internetworking

The home agent puts the whole IP datagram into an outer IP datagram to forward an IP datagram to a care-of address. This is a form of encapsulation, just as placing an IP header in front of a TCP segment encapsulates the TCP segment in an IP datagram. For mobile IP, three options for encapsulation are permitted –

- (i) IP-within-IP encapsulation
- (ii) Minimal encapsulation
- (iii) Generic routing encapsulation (GRE).

Q.50. Explain mobile IP registration process. How mobile TCP different from normal TCP ? (R.G.P.V., Dec. 2012)

Ans. Mobile IP Registration Process – Refer the ans. of Q.49.

Difference – Normal TCP reacts with slow start if acknowledgements are missing, which does not help in the case of transmission errors over wireless links and which does not really help during handover. Whereas mobile TCP does not use slow start, thus, M-TCP requires a bandwidth manager to implement fair sharing over the wireless link. As TCP has been designed for fixed networks with fixed end systems, the probable reason for a packet loss in this system is congestion at a node. Whereas in M-TCP, congestion is not

the reason for packet loss. Normal TCP relies on the congestion window to regulate transmission to the receiver, while M-TCP relies on the SH, instead of congestion window, to regulate transmission to the receiver.

Q.51. Explain the following with respect to mobile IPs –

(i) IP packet delivery (ii) Agent discovery (iii) Registration.

(R.G.P.V., Nov. 2019)

Ans. Refer the ans. of Q.49.

Q.52. Briefly explain applications of mobile IP. (R.G.P.V., Dec. 2015)

Ans. Some applications of mobile IP are given below –

(i) To support seamless and continuous internet connectivity mobile IP was designed.

(ii) Mobile user's point of attachment changes dynamically and all connections are automatically maintained despite the change.

(iii) User's internet connection is terminated each time the user moves and a new connection is initiated when the user dials back in.

Q.53. Explain deregistration process in mobile IP. (R.G.P.V., Dec. 2015)

Ans. It has to deregister with its home agent (HA) when a mobile node wants to drop its care-of-address. It achieves this by sending a registration request with the lifetime set to zero. If lifetime becomes zero, there is no requirement for deregistering with the FA as registration automatically expires. Therefore, the old foreign network does not know the new care-of-address of the mobile node when the mobile node visits a new network. Therefore, the packets are already forwarded through the HA true old FA of the mobile node are lost.

AD-HOC NETWORK ROUTING – AD-HOC NETWORK ROUTING V/S TRADITIONAL IP ROUTING, TYPES OF ROUTING PROTOCOLS, EXAMPLES – AODV, DSDV, DSR, ZRP, ETC.

Q.54. Write a short note on ad-hoc routing protocol. (R.G.P.V., June 2005)

Or

Explain how routing is performed on ad-hoc networks.

(R.G.P.V., Dec. 2013, 2016)

Ans. We have seen how to do routing when the hosts are mobile but the routers are fixed. However, what happens, in case when router themselves are mobile. Among the possibilities are –

- (i) Military vehicles on a battlefield with no existing infrastructure.
- (ii) A fleet of ships at sea.
- (iii) Emergency workers at an earthquake that destroyed the infrastructure.

(iv) A gathering of people with notebook computers.

In all the above cases, and others, each node consists of a router and a host, usually on the same computer. Networks of nodes that just happen to be near each other are called **ad-hoc networks** or **MANETs (mobile ad-hoc NETWORKs)**. Let us now examine them briefly. More information can be found in (Perkins, 2001).

What makes ad-hoc networks different from wired networks is that all the usual rules about fixed topologies, fixed and known neighbours, fixed relationship between IP address and location, and more are suddenly tossed out the window. Routers can come and go or appear in new places at the drop of a bit. With a wired network, if a router has a valid path to some destination, that path continues to be valid indefinitely (barring a failure somewhere in the system). With an ad-hoc network, the topology may be changing all the time, so desirability and even validity of paths can change spontaneously, without warning. Needless to say, these circumstances make routing in ad-hoc networks quite different from routing in their fixed counterparts.

A variety of routing algorithms for ad-hoc networks have been proposed. One of the more interesting ones is the **AODV (Ad-hoc On-demand Distance Vector)** routing algorithm (Perkins and Royer, 1999). It is a distant relative of the Bellman-Ford distance vector algorithm but adapted to work in a mobile environment and takes into account the limited bandwidth and low battery life found in this environment. Another unusual characteristic is that it is an on-demand algorithm, that is, it determines a route to some destination only when somebody wants to send a packet to that destination.

Q.55. Explain ad-hoc network in terms of characteristic and performance issue. (R.G.P.V., Dec. 2014)

Ans. Refer the ans. of Q.13 and Q.54.

Q.56. Explain routing in ad-hoc networks with suitable example.

Ans. In wireless networks, the infrastructure support a base station always reaches all mobile nodes, this is not always the case in an ad-hoc network. A destination node might be out of range of a source node transmitting packets. Routing is needed to find a path between source and destination and to forward the packets appropriately. In wireless networks using an infrastructure, cells have been defined. Within a cell, the base station can reach all mobile nodes without routing via a broadcast. In the case of ad-hoc networks, each node must be able to forward data for other nodes.

Fig. 3.40 shows a simple example of an ad-hoc network. At a certain time t_1 the network topology might look as given on the left side of the figure. Five nodes, N_1 to N_5 , are connected depending on the current transmission characteristics between them. In the network shown in fig. 3.40, N_4 can receive N_1 over a good link, but N_1 receives N_4 only via a weak link. Links do not

necessarily have the same characteristics in both directions. The reason for this are, e.g. different antenna characteristics or transmit power. N_1 cannot receive N_2 at all, N_2 receives a signal from N_1 .

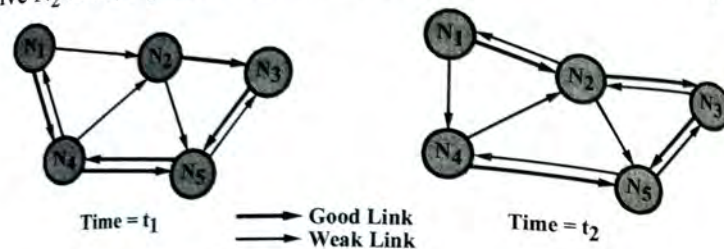


Fig. 3.40 Example Ad-hoc Network

This situation can change quite fast as the snapshot at t_2 shows. N_1 cannot receive N_4 any longer, N_4 receives N_1 only via a weak link. But now N_1 has an asymmetric but bi-directional link to N_2 that did not exist before.

Q.57. Write down the differences between Ad-hoc wireless network routing and wired network routing (traditional routing).

Ans. Differences between ad-hoc wireless networks routing and wired networks routing are as follows –

(i) Interference – In wired networks connections exist only where a wire exists, and connections are planned by network administrators. But in ad-hoc wireless networks, links come and go depending on transmission characteristics, one transmission might interfere with another, and nodes might overhear the transmissions of other nodes. Interference originates new problems by unplanned links between nodes i.e., if two close by nodes forward two transmissions, they might interfere and destroy each other. Alternatively, interference might also help routing. A node can learn the topology with the help of packets it has overheard.

(ii) Asymmetric Links – Node A receives a signal from node B. But this does not tell us anything about the connection quality in reverse. Node B might receive nothing, have a weak link, or even have a better link than the reverse direction. Routing information collected for one direction is of almost no use for the other direction. Although, many routing algorithms for wired networks rely on a symmetric scenario.

(iii) Redundant Links – Wired networks have redundant connections to survive link failures. Although, there is only some redundancy in wired networks, which additionally, are controlled by a network administrator. In ad-hoc networks nobody controls redundancy, so there might be many redundant connections up to the extreme of a completely meshed topology. In wired networks, routing algorithms can manage some redundancy, but a high redundancy may cause a large computational overhead for routing table updates.

(iv) Dynamic Topology – The greatest problem for routing arises from the highly dynamic topology. The mobile nodes might move or medium snapshots are valid only for a very short period of time. In ad-hoc networks, routing tables must somehow reflect these frequent changes in topology, and routing algorithms have to be adapted. Routing algorithms used in wired networks would either react much too slowly or generate too many updates to reflect all changes in topology. For example, routing table updates in fixed networks take place every 30 seconds. This updating frequency might be too low to be useful for ad-hoc networks. Some algorithms rely on a complete picture of the whole network. While this works in wired networks where changes are rare, it fails completely in ad-hoc networks. The topology changes during the distribution of the current snapshot of the network, rendering the snapshot useless.

Q.58. Why is routing in multi-hop ad-hoc network complicated? What are the special challenges?
(R.G.P.V., Dec. 2011)

Ans. Routing is complicated because of frequent topology changes, different capabilities of the nodes, varying propagation characteristics. Furthermore, no central instance can support routing.

There are many challenges involved, which are given below –

(i) Robustness – In ad-hoc wireless networks, network link failures are quite common because of the mobility of the nodes. Hence, the data packets transmitted by the source can be dropped, which gives in a low packet delivery ratio. Thus, a multicast routing protocol must be robust sufficient to control the mobility of the nodes and obtain a high packet delivery ratio.

(ii) Control Overhead – The exchange of control packets is needed to keep track of the members in a multicast group. This consumes a sufficient amount of bandwidth. As bandwidth is limited in ad hoc networks, the design of a multicast protocol should ensure that the total number of control packets sent for handling the multicast group is kept to a minimum.

(iii) Dependency on the Unicast Routing Protocol – It is difficult for the multicast protocol to work in heterogeneous networks, if a multicast routing protocol needs the support of a particular routing protocol. Thus, it is desirable if the multicast routing protocol is independent of any specific unicast routing protocol.

(iv) Efficiency – In an ad-hoc network environment, in which the bandwidth is limited, the multicast protocol efficiency is very important. The efficiency of multicast can be defined as the ratio of the total number of data packets received by the receivers to the total number of packets sent in the network.

(v) Resource Management – Ad-hoc networks consist of a group of mobile nodes, with each node having limited battery power and memory.

An ad-hoc multicast routing protocol must use less power by decreasing the number of packets transmissions. It should use less state information to decrease memory usage.

(vi) **Quality of Source (QoS)** – Ad-hoc networks are used in military/strategic applications. Thus, provisioning quality of service (QoS) is an issue in ad-hoc multicast routing protocols. The main parameters which are taken into consideration for giving the needed quality of service are throughput, delay, delay jitter, and reliability.

Q.59. What are the differences between AODV and the standard distance vector algorithm ?

Ans. The AODV routing protocol uses an on-demand approach for finding routes, that is, a route is established only when it is required by a source node for transmitting data packets. It employs destination sequence numbers to identify the most recent path. The major difference between AODV and dynamic source routing (DSR) stems out from the fact that DSR uses source routing in which a data packet carries the complete path to be traversed. However, in AODV, the source node and the intermediate nodes store the next-hop information corresponding to each flow for data packet transmission. In an on-demand routing protocol, the source node floods the *RouteRequest* packet in the network when a route is not available for the desired destination. It may obtain multiple routes to different destinations from a single *RouteRequest*. The major difference between AODV and other on-demand routing protocols is that it uses a destination sequence number (DestSeqNum) to determine an up-to-date path to the destination. A node updates its path information only if the *DestSeqNum* of the current packet received is greater or equal than the last *DestSeqNum* stored at the node with smaller hopcount.

Q.60. Explain the routing mechanism of destination sequence distance vector.

Ans. Destination sequence distance vector (DSDV) routing is an enhancement to distance vector routing for ad-hoc networks. In wired networks, distance vector routing is used as routing information protocol (RIP). It performs extremely poorly with certain network changes due to the count-to-infinity problem. Each node exchanges its neighbour table periodically with its neighbours. Changes at one node in the network propagate slowly through the network (step-by-step with every exchange). The strategies to avoid this problem which are employed in fixed networks do not help in the wireless ad-hoc networks, due to the rapidly changing topology. This might create loops or unreachable regions within the network.

Destination sequence distance vector now adds two things to the distance vector algorithm –

(i) **Damping** – Transient changes in topology that are of short duration should not destabilize the routing mechanisms. Advertisements containing

changes in the topology currently stored are therefore not disseminated further. If these changes are probably unstable, then a node waits with dissemination. Waiting time depends on the time between the first and the best announcement of a path to a certain destination.

The routing table for N_1 in fig. 3.40 would be as shown in table 3.2.

Table 3.2 Routing Table for DSDV

Destination	Next Hop	Sequence No.	Metric	Install Time
N_1	N_1	$S_1 - 321$	0	$T_4 - 001$
N_2	N_2	$S_2 - 218$	1	$T_4 - 001$
N_3	N_2	$S_3 - 043$	2	$T_4 - 002$
N_4	N_4	$S_4 - 092$	1	$T_4 - 001$
N_5	N_4	$S_5 - 163$	2	$T_4 - 002$

In each node N_1 stores the next hop toward this node, the metric, the sequence number of the last advertisement for this node, and the time at which the path has been installed first. The table contains flags and a settling time helping to decide when the path may be considered stable. Router advertisements from N_1 now contain data from the first, third, and fourth column-destination address, metric, and sequence number.

Q.61. Explain the routing mechanism of dynamic source routing. Describe the advantages of dynamic source routing over other routing approach in view of fixed networks.

(R.G.P.V., Dec. 2008)

Or

How does dynamic source routing handle routing ? (R.G.P.V., Dec. 2011)

Ans. Dynamic source routing (DSR) splits the task of routing into two problems –

(i) **Route Discovery** – A node attempts to discover a route to a destination if it contains something to send to this destination and at present there is no known route.

(ii) **Route Maintenance** – If a node is sending packets continuously through a route, it has to ensure that the route is held upright. A node finds an alternative as soon as it detects problems with the current route.

Dynamic source routing discards all periodic routing updates and works as follows. If a node wants to find a route, it broadcasts a route request with a unique identifier and the destination address as parameters. The following is done by any node that receives a route request –

(i) If the node identifies its own address as the destination, the request has reached its target.

(ii) If the node has already received the request, it drops the request packet.

(iii) Otherwise, the node adds its own address to a list of traversed hops in the packet and broadcasts this updated route request.

The route request collects a list of addresses representing a possible path on its way towards the destination by using this approach. As soon as the request reaches the destination, it can return the request packet containing the list to the receiver using this list in reverse order. One condition for this is that the links work bi-directionally. If this is not the case, and the destination node does not currently maintain a route back to the initiator of the request, it has to start a route discovery by itself. The destination may receive several lists containing different paths from the initiator. If could return the best path, the first path or several paths to offer the initiator a choice.

Applying route discovery to the example in fig. 3.40, for a route from N_1 to N_3 at time t_1 results in the following –

(i) N_1 broadcasts the request $((N_1), id = 42, target = N_3)$, N_2 and N_4 receive this request.

(ii) N_2 then broadcast $((N_1, N_2), id = 42, target = N_3)$, N_4 broadcast $((N_1, N_4), id = 42, target = N_3)$. N_3 and N_5 receive N_2 's broadcast, N_1 , N_2 and N_5 receive N_4 's broadcast.

(iii) N_3 recognize itself as target, N_5 broadcast $((N_1, N_2, N_5), id = 42, target = N_3)$. N_3 and N_4 receive N_5 's broadcast. N_1 , N_2 and N_5 drop N_4 's broadcast packet, because they all recognize an already received route request (and N_2 's broadcast reached N_5 before N_4 's did).

(iv) N_4 drops N_5 's broadcast, N_3 recognizes (N_1, N_2, N_5) as an alternate, but longer route.

(v) N_3 now has to return the path (N_1, N_2, N_3) to N_1 . This is simple assuming symmetric links working in both directions. N_3 can forward the information using the list in reverse order.

The assumption of bi-directional links holds for many ad-hoc networks. However, if links are not bi-directional, the scenario gets more complicated. The algorithm has to be applied again, in the reverse direction if the target does not maintain a current path to the source of the request.

(i) N_3 has to broadcast a route request $((N_3), id = 17 target = N_1)$. Only N_5 receives this request.

(ii) N_5 now broadcast $((N_3, N_5), id = 17, target = N_1)$, N_3 and N_4 receive the broadcast.

(iii) N_3 drops the request because it recognizes an already known id. N_4 broadcast $((N_3, N_5, N_4), id = 17 target = N_1)$, N_5 , N_2 , and N_1 receive the broadcast.

(iv) N_5 drops the request packet, N_1 recognizes itself as target, and N_2 broadcasts $((N_3, N_5, N_4, N_2), id 17, target = N_1)$. N_3 and N_5 receive N_2 's broadcast.

(v) N_3 and N_5 drop the request packet. Now N_3 holds the list for a path from N_1 to N_3 , (N_1, N_2, N_3) and N_1 known the path from N_3 to N_1 $(N_3,$

$N_5, N_4, N_1)$. But N_1 still does not know how to send data to N_3 . The only solution is to send the list (N_1, N_2, N_3) with the broadcasts initiated by N_3 in the reverse direction. This example shows clearly how much simpler routing can be if links are symmetrical.

Although dynamic source routing offers benefits compared to other algorithms by being much more bandwidth efficient, problems arise if the topology is highly dynamic and links are asymmetrical.

Q.62. Explain the zone routing protocol (ZRP) in detail.

Ans. The zone routing protocol is based on the concept of zones. A routing zone is defined for each node separately, and the zones of neighbouring nodes overlap. The routing zone has a radius ρ expressed in hops. The zone thus includes the nodes, whose distance from the node is at most ρ hops. An example routing zone is shown in fig. 3.41, where the routing zone of S includes the nodes A-I, but not K. In the illustrations, the radius is marked as a circle around the node. It should however be noted that the zone is defined in hops, not as a physical distance.

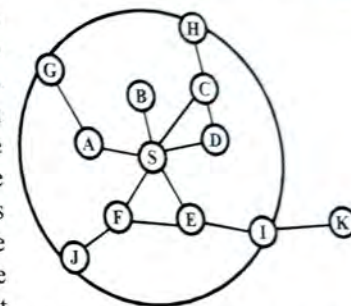


Fig. 3.41 Example Routing Zone with $\rho = 2$

The nodes of a zone are divided into peripheral nodes and interior nodes. Peripheral nodes are nodes whose minimum distance to the central node is exactly equal to the zone radius ρ . The nodes whose minimum distance is less than ρ are interior nodes. In fig. 3.41, the nodes A-F are interior nodes, the nodes G-J are peripheral nodes and the node K is outside the routing zone. Note that node H can be reached by two paths, one with length 2 and one with length 3 hops. The node is however within the zone, since the shortest path is less than or equal to the zone radius.

The number of nodes in the routing zone can be regulated by adjusting the transmission power of the nodes. Lowering the power reduces the number of nodes within direct reach and vice versa. The number of neighbouring nodes should be sufficient to provide adequate reachability and redundancy. On the other hand, a too large coverage results in many zone members and the update traffic becomes excessive. Further, large transmission coverage adds to the probability of local contention.

ZRP consist of three parts IARP proactive part, IERP reactive part of it and BRP used with IERP to reduce the query traffic.

(i) **IARP (Intra Zone Routing Protocol)** – The intra zone routing protocol (IARP) is a limited scope proactive routing protocol, which used to

support a primary global routing protocol. The routing zone radius shows the scope of the proactive part, the distance in hops that IARP route updates relayed. IARP's proactive tracking of local network connectivity provides support for route acquiring and route maintenance. First, routes to local nodes are immediately available, avoiding the traffic overhead and latency of a route discovery. Traditional proactive link state protocols modified to serve as an IARP by limiting link state updates to the scope of the link source's routing zone.

(ii) **IERP (Inter Zone Routing Protocol)** – The interzone routing protocol (IERP) is the global reactive routing component of the zone routing protocol (ZRP). IERP adapts existing reactive routing protocol implementations to take advantage of the known topology of each node's surrounding R-hop neighbourhood (routing zone), provided by the interzone routing protocol (IARP). The availability of routing zone routes allows IERP to suppress route queries for local destinations. When a global route discovery is required, the routing zone based border cast service used for efficiently guide route queries outward, rather than blindly relaying queries from neighbour to neighbour. Once a route discovered, IERP can use routing zones automatically to redirect data around failed links similarly, suboptimal route segments identified and traffic re-routed along shorter paths.

(iii) **BRP (Bordercast Resolution Protocol)** – The bordercast resolution protocol (BRP) provides the bordercasting packet delivery service. The BRP uses a map of an extended routing zone, provided by the local proactive intrazone routing protocol (IARP), to construct bordercast (multicast) trees along which query packets are directed. (Within the context of the hybrid ZRP, the BRP used to guide the route requests of the global reactive interzone routing protocol (IERP). The BRP uses special query control mechanisms to steer route requests away from areas of the network that have already covered by the query. The combination of multicasting and zone based query control makes bordercasting an efficient and tunable service that is more suitable than flood searching for network probing applications like route discovery. The bordercast resolution protocol (BRP) is a packet delivery service, not a full featured routing protocol. Bordercasting enabled by local proactive intrazone routing protocol (IARP) and supports global reactive interzone routing protocol (IERP).



UNIT

4

MOBILE TRANSPORT LAYER

UNSUITABILITY OF TRADITIONAL TCP, I-TCP, S-TCP, M-TCP

Q.1. What is the problems with traditional TCP in wireless environments ?

Ans. Problems with traditional TCP in wireless environments are as follows –

(i) Mobility itself can cause packet loss. There are many situations where a soft handover from one access point to another is not possible for a mobile end-system.

(ii) Error rates on wireless links are orders of magnitude higher compared to fixed fiber or copper links. This makes compensation for packet loss by TCP quite difficult.

(iii) Slow start mechanism in fixed networks decreases the efficiency of TCP if used with mobile receivers or senders.

(iv) Standard TCP reacts with slow start if acknowledgements are missing, which does not help in the case of transmission errors over wireless links and which does not really help during handover. This behaviour results in a severe performance degradation of an unchanged TCP if used together with wireless links or mobile nodes.

Q.2. What is the reaction of standard TCP in case of packet loss ? In what situation does this reaction make sense and why it is quite problematic in the case of wireless networks and mobility ? (R.G.P.V., June 2005)

Ans. TCP has been developed for fixed networks with fixed end-systems. Data transmission occurs using network adapters, fiber optics, copper wires, special hardware for routers, etc. Typically, this hardware works without introducing transmission errors. In case the software is sufficiently mature, it will not drop packets or flip bits, therefore if a packet on its way from a sender to a receiver is lost in a fixed network, it is not due to hardware, or software errors. In a fixed network, the probable reason for a packet loss is a temporary overload some point in the transmission path, i.e., a state of congestion at a node.

Congestion may occur from time to time even in carefully designed networks. Packet buffers of a router are filled and the router cannot forward the packets fast enough since the sum of the input rates of packets destined for one output link is higher as compared to the capacity of the output link. In this situation, the only thing a router can do, is to drop packets. A dropped packet is lost for the transmission, and the receiver sees a gap in the packet stream. Now, the receiver does not directly tell the sender which packet is missing, but continues to acknowledge all in-sequence packets up to the missing one.

After noticing the missing acknowledgement for the lost packet, the sender assumes a packet loss owing to congestion. Now, retransmitting the missing packet and continuing at full sending rate would be unwise, as this might only enhance the congestion. Although it is not guaranteed that all packets of the TCP connection take the same way through the network, this assumption holds for most of the packets. TCP slows down the transmission rate dramatically to mitigate congestion. All other TCP connections experiencing the same congestion do exactly the same, therefore the congestion is soon resolved. Using UDP is not a solution, since the throughput is higher than a TCP connection just at the starting. As soon as every one uses UDP, this advantage disappears. After that, congesting is standard and data transmission quality is unpredictable. TCP guarantees sharing of the bandwidth even under heavy load.

The reaction of the TCP to a missing acknowledgement is very drastic, but it is essential to get rid of congestion rapidly. The behaviour TCP shows after the detection of congestion is known as *slow start*. The sender always computes a congestion window for a receiver. Start size of the congestion window is one segment (TCP packet). The sender transmits one packet and waits for acknowledgement. When this acknowledgement comes, the sender increases the congestion window one, now transmitting two packets (congestion window = 2). Again, the sender adds 2 to the congestion window after arrival of the two corresponding acknowledgements, one for each of the acknowledgements. Then, the congestion window equals 4. Everytime the acknowledgements come back, this scheme doubles the congestion window, which takes one round trip time (RTT). This is known as the exponential growth of the congestion window in the slow start mechanism.

It is very risky to double the congestion window each time since the steps might become very large. At the congestion threshold, the exponential growth stops. As soon as the congestion window reaches the congestion threshold, further increase of the transmission rate is only linear by adding one to the congestion window each time the acknowledgements return.

The linear increase continues until the sender detects a gap in transmitted data due to continuous acknowledgements for the same packet, or until a time-

out at the sender takes place because of a missing acknowledgement. In either case, the sender sets the congestion threshold to half of the current congestion window. Congestion window itself is set to one segment and the sender begins transmitting a single segment. Exponential growth begins once more up to the new congestion threshold, then the window increases in linear manner.

A reduction of the congestion threshold occurs due to two things. One is a sender receiving continuous acknowledgements for the same packet. This informs the sender of two things. One is that the receiver got all packets upto the acknowledged packet in sequence. In TCP, a receiver transmits acknowledgements only when it receives any packets from the sender. Receiving acknowledgements from a receiver also depicts that the receiver continuously receives something from the sender. In the packet stream, the gap is not because of severe congestion, but a simple packet loss owing to a transmission error. Now, the sender can retransmit the missing packets before the timer expires. This behaviour is known as fast retransmit. Since the receipt of acknowledgements depicts that there is no congestion to justify a slow start, the sender can continue with the current congestion window. The sender performs a fast recovery from the packet loss. By using this mechanism, the efficiency of TCP can be improved dramatically.

A time-out due to a missing acknowledgement is another reason for activating slow-start. TCP using fast retransmit/fast recovery interprets this congestion in the network and activates the slow start mechanism.

The slow-start mechanism drastically decreases the efficiency of TCP if used together with mobile receivers or senders because slow start is used under the wrong assumptions. TCP concludes a congestion situation from a missing acknowledgement. While this may also occur in networks with mobile and wireless end-systems, it is not the main reason for packet loss.

Packet loss is much more common on wireless links and cannot always be compensated for by layer 2 retransmissions (ARQ) or error correction (FEC). Layer 2 faces the problem of transmitting the same packet twice over a bad link. Detecting these duplicates on layer 2 is not an alternative, because more and more connections use end-to-end encryption, making it impossible to look at the packet.

Mobility itself can cause packet loss. In many situations, a soft handover from one access point to another is not possible for a mobile end-system. As an example, when using mobile IP, there could still be some packets in transit to the old foreign agent while the mobile node moves to the new foreign agent. Old foreign agent may not be able to forward those packets to the new foreign agent or even buffer the packets if disconnection of the mobile node takes very long. This packet loss has nothing to do with wireless access but is caused by the problems of rerouting traffic.

The fundamental design problem in TCP is that the TCP mechanism detecting missing acknowledgements via time-outs and concluding packet loss owing to congestion cannot distinguish between the different causes and an error control mechanism (missing acknowledgement owing to a transmission error) is misused for congestion control (missing acknowledgement owing to network overload). In both cases, packets are lost (either because of invalid checksums or to dropping in routers). The reasons are totally different, but TCP cannot differentiate between these two different reasons. Standard TCP reacts with slow start if acknowledgements are missing, that does not help in the case of transmission errors over wireless links and that does not really help during handover. This behaviour causes a severe performance degradation of an unaltered TCP if used together with wireless links or mobile nodes.

Q.3. Show the interaction of mobile IP with standard TCP. Draw the packet flow from a fixed host to a mobile host via a foreign agent, then a handover takes place. What are following action of mobile IP and how does TCP react ? (R.G.P.V., June 2005, May/June 2006, June 2011)

Or

Why indirect TCP is developed ? How packet loss is handled in case of fixed network to mobile host and mobile host to fixed networks. How does I-TCP transaction protocol work ? (R.G.P.V., Dec. 2012)

Or

Explain the term indirect TCP.

(R.G.P.V., Dec. 2014)

Ans. Indirect TCP (I-TCP) is developed by keeping two things in mind. One is that TCP performs poorly together with wireless links, the other is that TCP within the fixed network cannot be altered. I-TCP splits a TCP connection into a fixed part and a wireless part. Fig. 4.1 illustrates an example with a mobile host connected via a wireless link and an access point to the 'wired' Internet where the correspondent host resides. The correspondent node could also use wireless access. Then, the following would also be applied to the access link of the correspondent host.

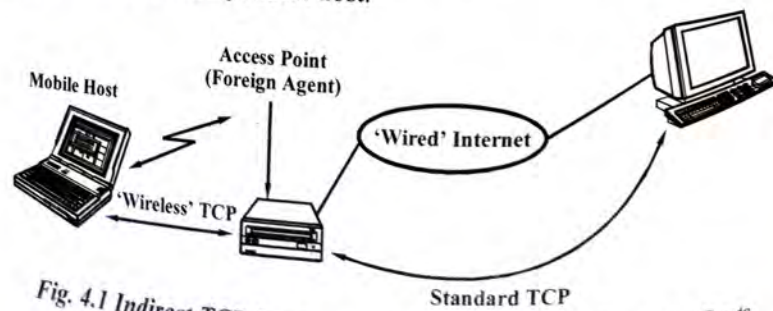


Fig. 4.1 Indirect TCP Segments a TCP Connection into two Parts

Standard TCP is employed between the fixed computer and the access point. No computer in the Internet identifies any alterations to TCP. Now, instead of the mobile host, the access point terminates the standard TCP connection, acting as a proxy. Then, the access point is seen as the mobile host for the fixed host and as the fixed host for the mobile host. A special TCP, adapted to wireless links, is used between the access point and the mobile host. However, it is not required to change TCP for the wireless link. Even an unaltered TCP can benefit from the much shorter round trip time, starting retransmission much faster. A good place for splitting the connection between mobile host and correspondent host is at the foreign agent of mobile IP. Foreign agent controls the mobility of the mobile host anyway. When the mobile host moves on, it can also handover the connection to the next foreign agent. However, separating the TCP connections can also be imagined at a special server, for example, at the entry point to a mobile phone network.

In the fixed network, the correspondent host does not notice the wireless link or the segmentation of the connection. Foreign agent acts like a proxy and relays all data in both directions. When the correspondent host transmits a packet, foreign agent acknowledges this packet and tries to forward the packet to the mobile host. When mobile host receives the packet, it acknowledges the packet. However, only the foreign agent uses this acknowledgement. In case a packet is lost on the wireless link because of a transmission error, the correspondent host would not notice it. In such a case, to maintain reliable data transport, the foreign agent attempts to retransmit this packet locally. Likewise, when the mobile host transmits a packet, the foreign agent acknowledges this packet and tries to forward it to the correspondent host. If the packet is lost on the wireless link, the mobile hosts notice it much faster owing to the lower round trip time and can directly retransmit the packet. In the wired network, packet loss is now handled by the foreign agent.

As soon as a handover occurs, I-TCP needs several actions. As fig. 4.2 illustrates, not only the packets have to be redirected using, e.g., mobile IP. In the example shown, the access point works like a proxy buffering packets for retransmission. After the handover, the old proxy must forward buffered data to the new proxy since it has already acknowledged the data. After registration with the new foreign agent, this new foreign agent can inform the old one about its location to enable packet forwarding. Besides buffer content, the sockets of the proxy must migrate to the new foreign agent located in the access point. The socket shows the current state of the TCP connection, i.e., sequence number, addresses, ports etc. No new connection can be established for the mobile host, and the correspondent host must not notice any changes in connection state.

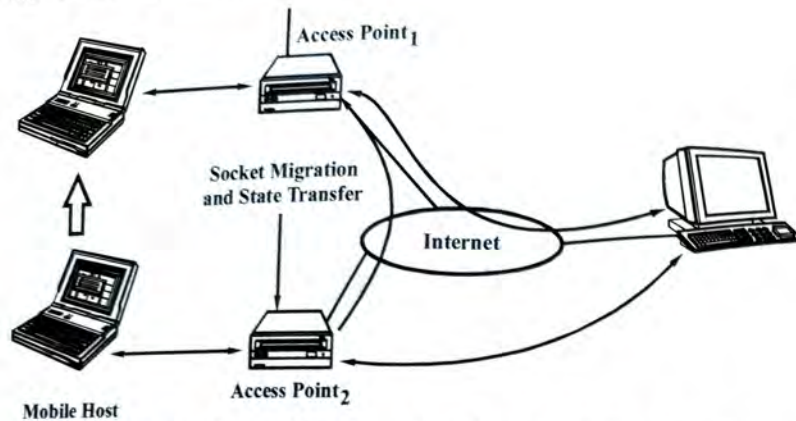


Fig. 4.2 Socket and State Migration after Handover of a Mobile Host

Q.4. How handover takes place in indirect TCP ? (R.G.P.V., June 2010)

Ans. Refer the ans. of Q.3.

Q.5. Write down the advantages and disadvantages of indirect TCP. (R.G.P.V., June 2004, Dec. 2009)

Ans. There are following advantages of I-TCP –

(i) It is dangerous to introduce new mechanisms into a huge network such as the Internet. However, new mechanisms are needed to improve TCP performance but with I-TCP only between the mobile host and the foreign agent. Different solution can be tested or used at the same time without jeopardizing the stability of the Internet. Furthermore, optimizing of these new mechanisms is quite simple because they only cover one single hop.

(ii) Due to the strict partitioning into two connections, transmission errors on the wireless link, i.e., lost packets, cannot propagate into the fixed network. Without partitioning, retransmission of lost packets would take place between mobile host the correspondent host across the whole network. Now only packets in sequence, without gaps leave the foreign agent.

(iii) I-TCP does not require any changes in the TCP protocol as used by the hosts in the fixed network or other hosts in a wireless network that do not use this optimization.

(iv) Partitioning into two connections also allows the use of a different transport layer protocol between the foreign agent and the mobile host or the use of compressed headers etc. The foreign agent can now act as a gateway to translate between the different protocols.

The idea of segmentation in I-TCP also comes with some disadvantages–

(i) The loss of the end-to-end semantics of TCP might cause problems if the foreign agent partitioning the TCP connection crashes. If a

sender receives an acknowledgement, it assumes that the receiver got the packet. Receiving an acknowledgement now only means (for the mobile host and a correspondent host) that the foreign agent received the packet. The access node may also crash applications running on the correspondent node assuming reliable end-to-end delivery.

(ii) The foreign agent must be a trusted entity because of TCP connections end at this point. If users apply end-to-end encryption, e.g., according to RFC 2401, the foreign agent has to be integrated into all security mechanisms.

Q.6. Write short note on snooping TCP.

(R.G.P.V., June 2011, Dec. 2013)

Or

Explain the term snooping TCP.

(R.G.P.V., Dec. 2014)

Ans. A disadvantage of I-TCP is that the segmentation of the single TCP connection into two TCP connections. This loses the original end-to-end TCP semantic. The following TCP enhancement works completely transparently and leaves the TCP end-to-end connection intact. The main function of the enhancement is to buffer data close to the mobile host to perform fast local retransmission in case of packet loss.

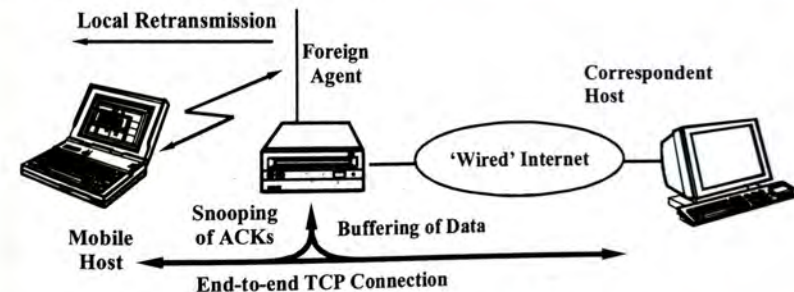


Fig. 4.3 Snooping TCP

In this approach, the foreign agent buffers all packets with destination mobile host and additionally 'snoops' the packet flow in both directions to recognize acknowledgements. The reason for buffering packets toward the mobile node is to enable the foreign agent to perform a local retransmission in case of packet loss on the wireless link. The foreign agent buffers every packet until it receives an acknowledgement from the mobile host within a certain amount of time, either the packet or the acknowledgement has been lost. Now the foreign agent retransmits the packet directly from the buffer, performing a much faster retransmission compared to the correspondent host. The time out for acknowledgements can be much shorter, because it reflects only the delay of one hop plus processing time.

The foreign agent must not acknowledge data to the correspondent host for transparency. This would make the correspondent host believe that the mobile host has received the data and would violate the end-to-end semantic in case of a foreign agent failure. However, the foreign agent can filter the duplicate acknowledgements to avoid unnecessary retransmissions of data from the correspondent host. If the foreign agent now crashes, the time-out of the corresponding host still works and triggers a retransmission. The foreign agent may discard duplicates of packets already retransmitted locally and acknowledged by the mobile host. This avoids unnecessary traffic on the wireless link.

Data transfer from the mobile host with destination correspondent host works as follows. The foreign agent snoops into the packet stream to detect gaps in the sequence number of TCP. When foreign agent detects a missing packet, it returns a negative acknowledgement (NACK) to the mobile host. The mobile host can now immediately retransmit the missing packet. Reordering of packets is done automatically at the correspondent host by TCP.

Q.7. With advantages and disadvantages, explain indirect TCP. Explain snooping TCP.
(R.G.P.V., Nov. 2018)

Ans. Refer the ans. of Q.3, Q.5 and Q.6.

Q.8. Explain snooping TCP. What are its advantages and disadvantages ?
(R.G.P.V., June 2009, 2010)

Or

Describe snooping TCP. Write advantages and disadvantages of snooping TCP.
(R.G.P.V., Dec. 2008, 2017)

Ans. Snooping – Refer the ans. of Q.6.

Advantages and Disadvantages – There are following advantages of snooping TCP –

(i) As soon as the mobile host moves to another foreign agent, it does not require a handover of state. Suppose there might still be data in the buffer not transferred to the next foreign agent. All that occurs is a time-out at the correspondent host and retransmission of the packets, possibly already to the new care-of address.

(ii) End-to-end TCP semantic is preserved. No matter at what time the foreign agent crashes, neither the correspondent host nor the mobile host have an inconsistent view of the TCP connection as is possible with I-TCP. The approach automatically falls back to standard TCP if the enhancement stop working.

(iii) It is not important if the next foreign agent uses the enhancement or not. If not, the approach automatically falls back to the standard solution. This is one of the problems of I-TCP because the old foreign agent may have

already signaled the correct receipt of data via acknowledgements to the correspondent host and now has to transfer these packets to the mobile host via the new foreign agent.

(iv) Correspondent host does not require to be altered; most of the enhancements are in the foreign agent. Supporting only the packet stream from the correspondent host to the mobile host does not even need changes in the mobile host.

There are following disadvantages of snooping TCP –

(i) Applying negative acknowledgements between the foreign agent and the mobile host assumes additional mechanisms on the mobile host. This method is no longer transparent for mobile hosts.

(ii) If certain encryption schemes are applied end-to-end between the correspondent host and mobile host, all efforts for snooping and buffering data may be useless. Using IP encapsulation security payload, the TCP protocol header will be encrypted-snooping on the sequence number will no longer work. Retransmitting data from the foreign agent may not work since several security schemes prevent replay attacks – retransmitting data from the foreign agent may be misinterpreted as replay. If encryption is used above the transport layer, snooping TCP can be used.

(iii) Snooping TCP does not isolate the behaviour of the wireless link as well as I-TCP.

Q.9. Explain the agent advertisement method of mobile IP. What are the advantages and disadvantages of snooping TCP ? (R.G.P.V., Dec. 2012)

Ans. Refer the ans. of Q.49, (U-III) and Q.8.

Q.10. Write short note on mobile TCP. (R.G.P.V., Dec. 2008, 2013)

Or

Define mobile TCP with its objectives. (R.G.P.V., Dec. 2015)

Ans. Dropping packets due to a handover or higher bit error rates is not the only phenomenon of wireless links and mobility – the occurrence of lengthy and frequent disconnections is another problem. Quite often mobile users cannot connect at all.

A TCP sender tries to retransmit data controlled by a retransmission timer that doubles with each unsuccessful retransmission attempt, upto a maximum of one minute (the initial value depends on the round trip time). This means that the sender tries to retransmit an unacknowledged packet every minute and will give up after 12 retransmissions. No data is successfully transmitted for a period of one minute! The retransmission time-out is still valid and the sender has to wait. The sender also goes into slow-start because it assumes congestion.

The M-TCP (mobile TCP) approach has the same goals as I-TCP and snooping TCP – to prevent the sender window from shrinking if bit errors of disconnection but not congestion cause current problems. M-TCP is used to improve overall throughput, to lower the delay, to maintain end-to-end semantics of TCP, and to provide a more efficient handover.

M-TCP splits the TCP connection into two parts as I-TCP does. An unmodified TCP is used on the standard host-supervisory host (SH) connection, while an optimized TCP is used on the SH-MH connection. The supervisory host is responsible for exchanging data between both parts. The M-TCP approach assumes a relatively low bit error rate on the wireless link. Therefore, it does not perform caching / retransmission of data via the SH. If a packet is lost on the wireless link, it has to be retransmitted by the original sender. This maintains the TCP end-to-end semantics.

The SH monitors all packets sent to the MH and ACKs returned from the MH. If the SH does not receive an ACK for some time, it assumes that the MH is disconnected. It then chokes the sender by setting the sender's window size to 0. Setting the window size to 0 forces the sender to go into persistent mode, i.e., the state of the sender will not change no matter how long the receiver is disconnected. This means that the sender will not try to retransmit data. As soon as the SH (either the old SH or a new SH) detects connectivity again, it reopens the window of the sender to the old value. The sender can continue sending at full speed. This mechanism does not require changes to the sender's TCP.

Q.11. Write a brief notes on snooping TCP and mobile TCP.

(R.G.P.V., Dec. 2016)

Ans. Refer the ans. of Q.6 and Q.10.

Q.12. Differentiate between traditional TCP and mobile TCP. Also explain the working of indirect TCP.

(R.G.P.V., Nov. 2019)

Ans. Refer the ans. of Q.2, Q.10 and Q.3.

Q.13. Compare and contrast I-TCP, snooping TCP and mobile TCP.

(R.G.P.V., Nov. 2019)

Ans. Refer the ans. of Q.3, Q.6 and Q.10.

Q.14. Write the advantages of mobile TCP.

Ans. There are following advantages of mobile TCP –

- (i) M-TCP maintains the TCP end-to-end semantics. The SH does not transmit any ACK itself but forwards the ACKs from the MH.

- (ii) Because it does not buffer data in the SH as I-TCP does, it is not essential to forward buffers to a new SH. Lost packets will be automatically retransmitted to the new SH.

- (iii) When the MH is disconnected, it avoids useless retransmissions, slow starts or breaking connections by simply shrinking the sender's window to 0.

Q.15. What are the limitations of mobile TCP? (R.G.P.V., Dec. 2015)

Ans. There are following limitations of mobile TCP –

- (i) Since the SH does not act as proxy as in I-TCP, packet loss on the wireless link because of bit errors is propagated to the sender. M-TCP assumes low bit error rates, that is not always a valid assumption.

- (ii) A modified TCP on the wireless link not only needs modifications to the MH protocol software but also new network elements like the band width manager.

Q.16. Write short note on timeout freezing. (R.G.P.V., June 2004, 2008)

Ans. I-TCP, N-TCP, Snooping TCP approaches can handles short interruption of connection, for longer interruption of transmission. Examples are the use of mobile hosts in a car driving into a tunnel, which loses its connection to a satellite or a user moving into a cell with no capacity left over. In this case, the mobile phone system will interrupt the connection. Thus now the reaction of TCP would be disconnection after a timeout. The MAC layer knows the real reason for the interruption and does not assume congestion, as TCP would. The MAC layer can inform the TCP layer of an upcoming loss of connection or that the current interruption is not caused by congestion. TCP can now stop sending and 'freezes' the current state of its congestion window and further timers. If the MAC layer notices the upcoming interruption early enough, both the mobile and correspondent host can be informed. With a fast interruption of the wireless link, additional mechanisms in the access point are needed to inform the correspondent host of the reason for interruption. Otherwise, the correspondent host goes into slow that assuming congestion and finally breaks the connection.

As soon as the MAC layer detects connectivity again, it signals TCP that it can resume operation at exactly the same point where it had been forced to stop. For TCP time simply does not advance, so no timers expire.

The main advantage of this method is that it offers a way to resume TCP connections even after longer interruptions of the connection. It is independent of any other TCP mechanism, such as acknowledgements or sequence numbers, so it can be used together with encrypted data. However, this scheme has some severe disadvantages –

- (i) All mechanism rely on the capability of the MAC layer to detect future interruptions.
- (ii) Freezing the state of TCP does not help in case of some encryption schemes that use time-dependent random numbers.
- (iii) These schemes need resynchronization after interruption.

Q.17. Illustrate the TCP connection timeout freezing of transmission protocol.

Ans. Refer the ans. of Q.16.

NUMERICAL PROBLEMS

Prob.1. If 20 MHz of total spectrum is allocated for a duplex wireless cellular system and each simplex channel has 25 kHz RF bandwidth find –

(i) The no. of duplex channels

(ii) The total no. channel per cell site

If $N = 4$ cell reuse is used.

(R.G.P.V., Dec. 2012)

Sol. (i) The total bandwidth given is 20 MHz

$$\begin{aligned}\text{Channel bandwidth} &= 25 \text{ kHz} \times 2 \text{ simplex channels} \\ &= 50 \text{ kHz per duplex channel}\end{aligned}$$

The no. of duplex channels = $20 \text{ MHz} / 50 \text{ kHz} = 400$ channels *Ans.*

(ii) If $N = 4$, then the total number of available channels per cell is $400/4 = 100$. *Ans.*

WIRELESS CELLULAR NETWORKS – CELLULAR SYSTEM, CELLULAR NETWORKS V/S WLAN

Q.18. What are the unconventional applications of wireless networks?

(R.G.P.V., Dec. 2005)

Ans. Wireless networks find their applications in the following application environments –

(i) **Vehicles** – Today's car already comprise some, but tomorrow's car will comprise several wireless communication systems and mobility aware applications. Road conditions, weather reports, music, news, and other broadcast information are received via digital audio broadcasting (DAB) with 1.5 Mbit/s. A universal mobile telecommunication system (UMTS) phone might be available for personal communication offering voice and data connectivity

with 384 kbit/sec. For remote areas, satellite communication can be used, while the current position of the car is determined by via the global positioning system (GPS). Cars driving in the same area build a local ad-hoc network for the fast exchange of information in emergency situations or to help each other keep a safe distance. When an accident occurs, not only will the airbag be triggered, but the police and ambulance service will be informed via an emergency call to a service provider. This technology is already available in cars. In the future, cars will also inform other cars about accidents via the ad-hoc network to help them slow down in time, even before a driver can identify an accident. Buses, trucks, and trains are already transmitting maintenance and logistic information to their home base, that helps to improve organization (fleet management), and saves time and money.

A typical scenario for mobile communications with several wireless devices is shown in fig. 4.4. Networks with a fixed infrastructure such as cellular phones (GSM, UMTS) will be interconnected with trunked radio systems (TETRA) and wireless LANs. Satellite communication links can also be employed. The networks between cars and inside each car will more likely work in an ad-hoc fashion. Wireless pico networks inside a car can comprise personal digital assistants (PDA), laptops, or mobile phones, e.g., connected with each other using the bluetooth technology.

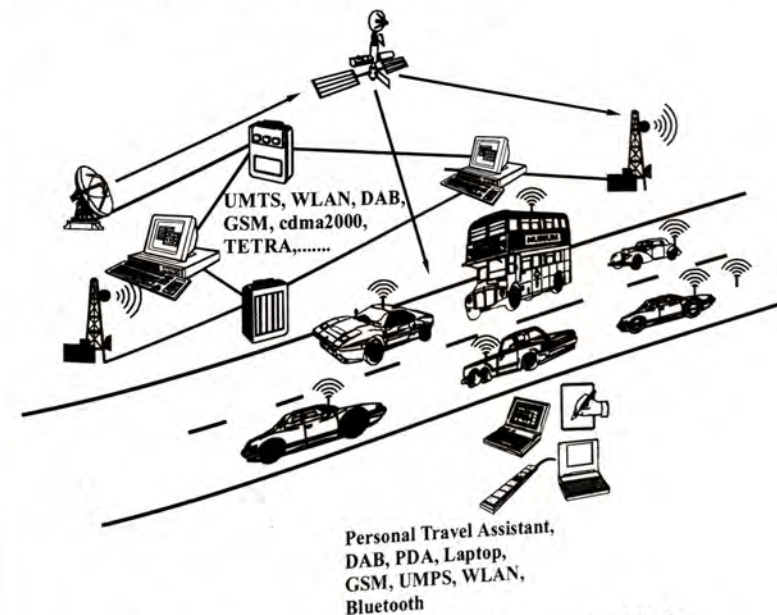


Fig. 4.4 Application of Wireless Networks in Vehicles

(ii) **Emergencies** – Wireless networks are the only means of communication in the case of natural disasters like hurricanes or earthquakes. In the worst cases, only decentralized, wireless ad-hoc networks survive. Breakdown of all cabling not only implies the failure of the standard wired telephone system, but also the crash of all mobile phone systems needing base stations.

(iii) **Business** – Now-a-days, a travelling salesman can require instant access to the database of the company to ensure that files on his or her laptop reflect the current situation, to enable the company to keep track of all activities of their travelling employees, to keep databases consistent etc. The laptop can be turned into a true mobile office with wireless access, but efficient and powerful synchronization mechanisms are required to ensure data consistency. Fig. 4.5 shows what may occur when employees try to communicate off base. At home, the laptop connects via a WLAN or LAN and DSL to the Internet. Leaving home needs a handover to another technology, for example, to an enhanced version of GSM, as soon as the WLAN coverage ends. Because of interference and other factors, data rates drop while cruising a higher speed. Gas stations may offer WLAN hot spots as well as gas. Trains already offer support for wireless connectivity. Many more handovers to different technologies might be essential before reaching the office. No matter when and where, mobile communications should always offer as good connectivity as possible to Internet, the Internet of the company, or the telephone network.

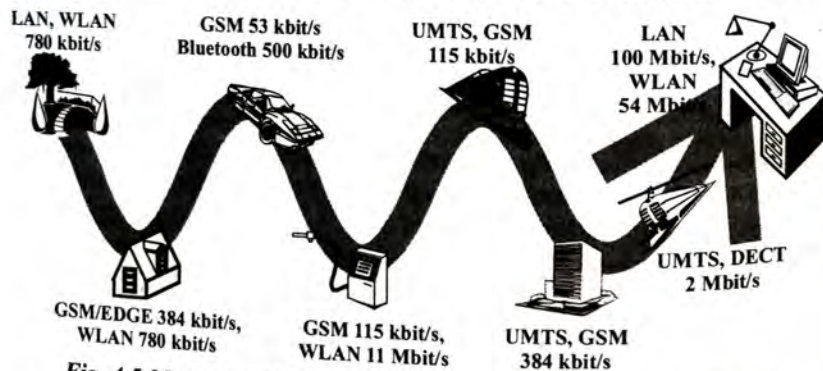


Fig. 4.5 Mobile and Wireless Services – Always Best Connected

(iv) **Replacement of Wired Networks** – Wireless networks can also be used to replace wired networks in some cases, for example, remote sensors, for tradeshows or in historic buildings. Owing to economic reasons, it is often not possible to wire remote sensors for weather forecasts, earthquake detection or to provide environmental information. In this situation, wireless connections, for example, via satellite, can help. Tradeshows require a highly dynamic

infrastructure, but cabling takes a long time and frequently proves to be too flexible. Several computer fairs use WLANs as a replacement for cabling. For wireless networks, other cases are sensors, computers, or information displays in historic buildings, where excess cabling can destroy valuable walls or floors. Wireless access points in a corner of the room can represent a solution.

(v) **Infotainment and More** – Wireless networks can provide up-to-date information at any appropriate location. The travel guide might tell you something about the history of a building downloading information about a concert in the building at the same evening via a local wireless network. We may select a seat, pay via electronic cash, and transmit this information to a service provider. Another growing field of wireless network applications lies in entertainment and games to enable, for example, ad-hoc gaming networks as soon as people meet to play together.

(vi) **Location Dependent Services** – Several services that might depend on the actual location can be distinguished –

(a) **Follow-on Services** – Using mobile computers, a follow-on service could offer, for example, the same desktop environment wherever you are in the world. All e-mail would automatically be forwarded and all changes to your desktop and documents would be stored at a central location at your company. If someone wanted to reach you using a multimedia conferencing system, this call would be forwarded to your current location.

(b) **Location Aware Services** – These services are very important in present environment.

(c) **Privacy** – Privacy is an important factor while providing two service classes listed above.

(d) **Information Services** – While walking around in a city you could always use your wireless travel guide to ‘pull’ information from a service. However, a service could also actively ‘push’ information on your travel guide.

(e) **Support Services** – Several small additional mechanisms can be integrated to support a mobile device. Intermediate results of calculations, state information or cache contents could ‘follow’ the mobile node through the fixed network. As soon as the mobile node reconnects, all information is available again. This helps in reducing access delay and traffic within the fixed network.

Q.19. Discuss the need of wireless system in contrast with the conventional communication system.
(R.G.P.V., June 2005)

Ans. The need of wireless system in contrast with the conventional communication system is given below –

(i) **Poor Service Performance** – A total of thirty three channels was allocated to three mobile telephone systems in the conventional mobile telephone system. These telephone systems are mobile telephone service (MTS), improved mobile telephone service (IMTS) MK systems and improved mobile telephone service (IMTS) MJ systems. Operating frequency of MTS is around 40 MHz and operating frequency for MJ is 150 MHz. Both MTS and MJ systems facilitate 11 channels. IMTS MK works at 450 MHz and provides 12 channels. An area 50 mile in diameter is covered by these 33 channels.

In year 1976, New York city also had 6 channels of MK system serving 225 customers, with another 1300 customers on a waiting list. New York city also has 6 channels of MJ system serving 320 customers, with another 2400 customers on a waiting list. During the busy hours, a high blocking probability is generated by the large number of customers. Although service performance of conventional mobile system was poor, the demand was still great. So a high-capacity system for mobile telephones was required.

(ii) **Limited Service Capability** – Fig. 4.6 shows a conventional mobile system is designed for use in autonomous geographic zones by choosing one or more channels from a particular frequency allocation. Usually, the communications coverage area of every zone is planned to be big as possible, it implies that the transmitted power should be as large as the federal specification permits. A call is started by the customer in single zone has to reinitiate the call when customer moving into a new zone due to the call will be dropped. There is no guarantee that a call will be completed without a handoff capability because this is an undesirable radio telephone system.

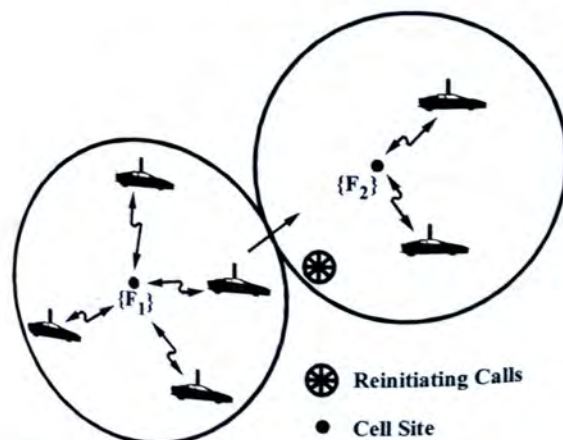


Fig. 4.6 Conventional Mobile System

Another limitation of the conventional system is that the number of active customer is limited to the many channels allocated to a specific zone of frequency.

(iii) **Inefficient Frequency Spectrum Utilization** – Frequency utilization measurement M_0 can be expressed by the maximum number of users which could be served by one channel at the busy hour in a conventional mobile telephone system.

$$M_0 = \frac{\text{Number of users}}{\text{Channel}} \quad (\text{for conventional systems})$$

$$\text{where, } M_0 = \begin{cases} 37 \text{ users/channel for MK system} \\ 53 \text{ users/channel for MJ system.} \end{cases}$$

Suppose that an average calling time is 1.76 min and we can apply the Erlang B model. Blocking probability can be determined as follows – use 6 channels, with each channel serving two different numbers of users. Then, offered load is expressed as –

$$A = \frac{\text{Average calling time (min)} \times \text{Total users}}{60 \text{ min}} \quad (\text{Erlangs})$$

For MK system,

$$A_1 = \frac{1.76 \times 37 \times 6}{60} = 6.51 \text{ erlangs}$$

For MJ system,

$$A_2 = \frac{1.76 \times 53 \times 6}{60} = 9.33 \text{ erlangs}$$

For offered loads $A_1 = 6.51$ and $A_2 = 9.33$, the obtained blocking probabilities are $B_1 = 30\%$ for MK system and $B_2 = 50\%$ for MJ system, respectively.

The blocking probability can be even higher, when the real average calling time is higher than 1.76 min. The frequency spectrum utilization measurement M_0 must be decreased for minimizing the blocking probability.

Therefore, conventional system does not utilize the frequency spectrum efficiently because each channel can only serve one customer at a time in the whole area.

Q.20. Explain wireless transmission impairments. (R.G.P.V., June 2004)

Ans. Wireless transmission has the following impairments –

(i) **Lower Security, Simpler to Attack** – Not only can portable device be stolen more easily, but the radio interface is also prone to the dangers of

eavesdropping. Wireless access must always include encryption, authentication, and other security mechanisms which must be efficient and simple to use.

(ii) **Low Bandwidth** – Although they are continuously increasing, transmission rates are still very low for wireless devices compared to desktop systems. Local wireless systems reach some Mbit/s while wide area systems only offer some 10 kbit/s. One task would involve adapting applications used with high-bandwidth connections to this new environment so that the user can continue using the same application when moving from the desktop outside the building. Researchers look for more efficient communication protocols with low overhead.

(iii) **Interference** – Radio transmission cannot be protected against interference using shielding as this is done in coaxial cable or shielded twisted pair. As an instance, electrical engines and lightning cause severe interference and result in higher loss rates for transmitted data or higher bit error rates respectively.

(iv) **Shared Medium** – Radio access is always realized via a shared medium. Since it is impossible to have a separate wire between a sender and each receiver, different competitors have to 'fight' for the medium. Although different medium access schemes have been developed, several questions are still unanswered, as example how to provide quality of service efficiently with different combinations of access, coding, and multiplexing schemes.

(v) **Regulations and Spectrum** – Frequencies have to be coordinated, and unfortunately, only a very limited amount of frequencies are available owing to technical and political reasons. One research topic involves determining how to use available frequencies more efficiently, e.g., by new modulation schemes or demand-driven multiplexing. New air interfaces, power aware ad-hoc networks, smart antennas, and software defined radios (SDR) are further improvements. The latter permit for software definable air interfaces but need high computing power.

(vi) **High Delays, Large Delay Variation** – The big variation in link characteristics is a serious problem for communication protocols used in today's Internet (TCP/IP). In wireless systems, delays of several seconds can take place, and links can be very asymmetrical (i.e., the links offer different service quality depending on the direction to and from the wireless device). Applications must be tolerant and use robust protocols.

(vii) **Ad-hoc Networking** – Wireless and mobile computing permits for spontaneous networking with prior set-up of an infrastructure. However, this originates several new questions for research – routing on the networking and application layer, service discovery, network scalability, reliability, and stability etc.

Q.21. What are the different configuration parameters to adapt the wireless environments ?
(R.G.P.V., Dec. 2008)

Ans. In wireless environment the main parameters to achieve the wireless environment is the transmission and reception of an antenna, the antenna radiates electromagnetic energy into the medium (usually air), and for reception, the antenna picks up electromagnetic waves from the surrounding medium. There are basically two types of configuration for wireless transmission – directional and omni-directional. For the directional configuration, the transmitting antenna puts out a focused electromagnetic beam; the transmitting and receiving antennas must therefore be carefully aligned. In the omni-directional case, the transmitted signal spreads out in all directions and can be received by many antennas. In general, the higher the frequency of a signal the more it is possible to focus it into a directional beam.

Three general ranges of frequencies are of interest in our discussion of wireless transmission. Frequencies in the range of about 2 GHz (gigahertz = 10^9 Hz) to 40 GHz are referred to as microwave frequencies. At these frequencies, highly directional beams are possible, and microwave is quite suitable for point-to-point transmission. Microwave is also used for satellite communication. Frequencies in the range of 30 MHz to 1 GHz are suitable for omni-directional applications. We will refer to this range as the broadcast radio range.

Another important frequency range, for local applications is the infrared portion spectrum. This cover, roughly from 3×10^{11} to 2×10^{14} Hz. Infrared is useful to local point-to-point and multipoint applications within confined areas such as a single room.

Q.22. Write short note on cellular network organization.

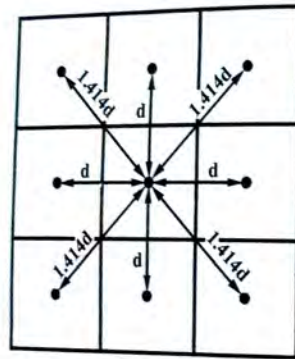
Or

What geometric shape is used in cellular system design ?

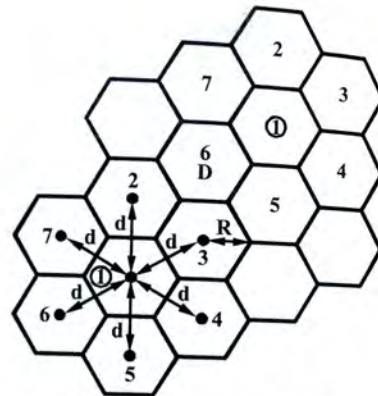
Ans. The essence of a cellular network is the use of multiple low-power transmitter, on the order of 100 W or less. Because the range of such transmitter is small, an area can be divided into cells, each one served by its own antenna. Each cell is allocated a band of frequencies and is served by a base station, consisting of transmitter, receiver and control unit. Adjacent cells are assigned different frequencies to avoid interference or crosstalk. However, cells sufficiently distant from each other can use the same frequency band.

The first design decision to make is the shape of cells to cover an area. A matrix of square cells would be the simplest layout to define [see fig. 4.7 (a)]. However, this geometry is not ideal. If the width of a square cell is d , then a cell had four neighbours at a distance d and four neighbours at a distance d

$\sqrt{2}d$. As a mobile user within a cell moves toward the cell's boundaries, it is best if all of the adjacent antennas are equidistant. This simplifies the task of determining when to switch the user to an adjacent antenna and which antenna to choose. A hexagonal pattern provides for equidistant antennas [see fig. 4.7 (b)]. The radius of a hexagon is defined to be the radius of the circle that circumscribes it. For a cell radius R , the distance between the cell center and each adjacent cell center is $d = \sqrt{3}R$.



(a) Square Pattern



(b) Hexagonal Pattern

Fig. 4.7 Cellular Geometries

Generally, a precise hexagonal pattern is not used. Variations from the ideal are due to topographical limitations, local signal propagation conditions, and practical limitation on siting antennas.

Q.23. Explain the operation of cellular system.

(R.G.P.V., Dec. 2006, June 2008)

Ans. The operation of cellular system can be explained as follows –

(i) **Mobile Unit Initialization** – The receiver of the mobile unit scans 21 set-up channels that are designated among the 416 channels if a user sitting in a car activates the receiver of the mobile unit. Then it chooses the strongest channel and locks on for a definite time. Because each site is allocated a different set-up channel, locking onto the strongest set-up channel generally means choosing the nearest cell site. This self-location method is user-independent and is used in the idle stage. As it eliminates the load on the transmission at the cell site to locate the mobile unit, this is the merit of self-location method. The demerit of the self-location method is that location information of idle mobile units does not appear at each cell site. Thus, the paging process is longer when the call originates from the land line to a mobile unit. The use of self-location methods is justified, since a large percentage of

calls generates at the mobile unit. The self-location procedure is repeated after 60 sec. When land-line generated calls increase in the future, a feature known as registration is employed.

(ii) **Mobile Originated Call** – The user places the known number into an generating register in the mobile unit, checks to observe that the number is correct and the send button is pushed. Then, on a selected set-up channel, a request for service is sent achieved from a self-location method. The cell site receives it and in directional cell sites, the best directive antenna is chosen for the voice channel to use. At the same time, a request is transmitted by the cell site to the mobile telephone switching office (MTSO) through a high-speed data link. A suitable voice channel is chosen by the MTSO for the call, and the cell site works on it by the best directive antenna to link the mobile unit. MTSO also connects the wire-line party by the telephone company zone office.

(iii) **Network Originated Call** – When a mobile unit number is dialed by a land-line party. The telephone company zone office identifies that the dialed number is mobile and forwards the call to the mobile telephone switching office (MTSO). Mobile telephone switching office (MTSO) transmits a paging message to certain cell sites dependent upon the mobile unit number and the search algorithm. Every cell site sends the page on its own set-up channel. On a strong set-up channel, the mobile unit identifies its own identification. Then, mobile unit locks onto it and responds to the cell site. Mobile unit also follows the instruction to tune to an allocated voice channel and initiate user alert.

(iv) **Call Termination** – If the mobile user switched off the transmitter, a particular signal (signalling tone) sends to the cell site and both sides free the voice channel. Through the strongest set-up channel, the mobile unit resumes controlling pages.

(v) **Handoff Procedure** – Two parties are on a voice channel, in the duration of the call. The reception becomes weak, at the time when mobile unit goes out of the coverage area of a particular cell site. A handoff is requested by the present cell site. Without either alerting the user or interrupting the call, the system switches the call to a new frequency channel in a new cell site. The call carries on as long as the user is talking. The handoff occurrences are not experienced by the user. The AMPS system first uses the **handoff**, then European system renames **handover** because the different meanings in American English and British English.

Q.24. Compare infrared and radio transmission and explain the architecture of a cellular system.

(R.G.P.V., Dec. 2012)

Ans. Comparison between Infrared and Radio Transmission – Table 4.1 gives a comparison among infrared, spread spectrum and radio transmission techniques –

Table 4.1

	Infrared		Spread Spectrum		Radio
	Diffused Infrared	Directed Beam Infrared	Frequency Hopping	Direct Sequence	Narrowband Microwave
Data Rate (Mbps)	1 to 4	1 to 10	1 to 3	2 to 20	10 to 20
Mobility	Stationary/mobile	Stationary with LOS	Mobile	Stationary/mobile	
Range (m)	15 to 60	25	30 to 100	30 to 250	10 to 40
Detectability	Negligible		Little		Some
Wavelength/frequency	λ : 800 to 900 nm		902 to 928 MHz 2.4 to 2.4835 GHz 5.725 to 5.85 GHz		902 to 928 MHz 5.2 to 5.775 GHz 18.825 to 19.205 GHz
Modulation technique	ASK		FSK	QPSK	FS/QPSK
Radiated power	-		<1W		25 mW
Access method	CSMA	Token Ring, CSMA	CSMA		Reservation ALOHA, CSMA
License required	No		No		Yes unless ISM

Architecture of a Cellular System – In a cellular system, the coverage area is partitioned into several smaller areas or cells. Each cell in the cellular system is served by a base station (BS) for radio coverage. A base station controller (BSC) is available to connect the BSs by using fixed links and it also connects to a mobile switching centre (MSC). MSC is a local switching exchange with additional features to handle mobility management requirements of a cellular system. The dynamic nature of terminal location information and subscription data are handled by the interaction of the MSC with appropriate databases that maintain subscriber data and location information. The MSC also interconnects with the PSTN because most of calls in a cellular mobile system either originate from or terminate at fixed network terminals or use fixed network like a transit network. The basic components of a typical cellular mobile system is shown in the fig. 4.8.

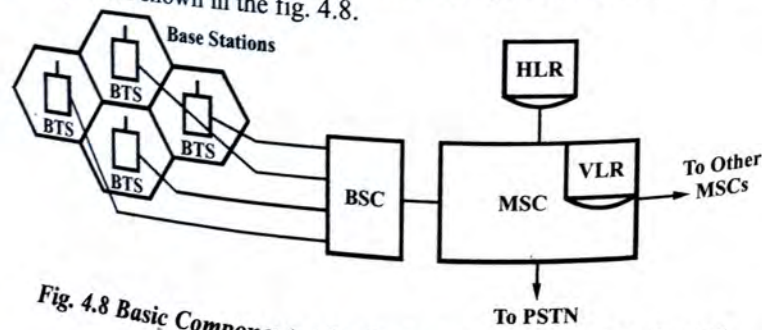


Fig. 4.8 Basic Components of a Typical Cellular Mobile System

Q.25. Describe the frequency reuse in cellular communication system. How it can be implemented to increase the capacity of the cellular communication system?

(R.G.P.V., Dec. 2008)

Ans. The concept of frequency reuse in a cellular system is illustrated in fig. 4.9. In each cell, by a hexagon, a group of radio "channel" frequencies is utilized. Cells designated with the same letter are assigned the same radio channels. For example, frequency group A is assigned to all cells (hexagons) designated by frequency group A in fig. 4.9. Thus the same frequencies are "reused" many times. The actual radio coverage depends on radio link parameters and propagation conditions. It is not hexagonal.

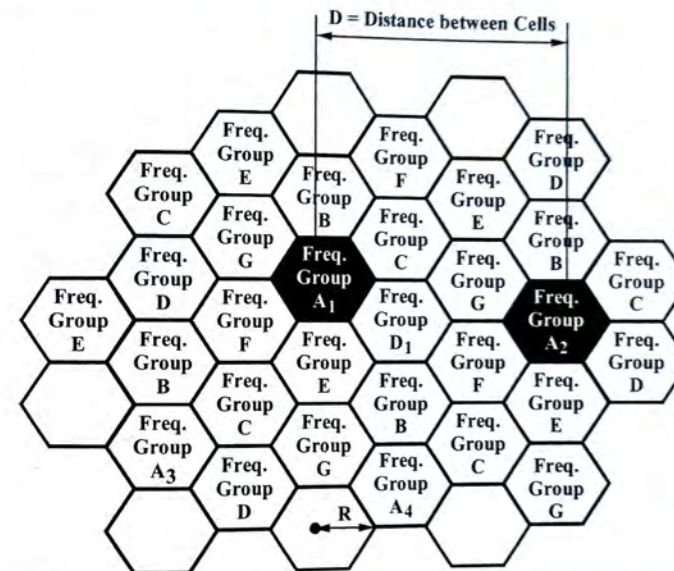


Fig. 4.9 Frequency Reuse in a Seven-cell-pattern Cellular System

The real power of the cellular idea is that interference is not related to the absolute distance between cells but to the ratio of the distance between same-frequency cells (D) to the cell radius (R).

Another attractive feature of the cellular concept is **cell splitting**. Larger cells can be easily reduced into smaller radius cells over a period of time through cell splitting. When the traffic reaches the point in a particular cell such that the existing allocation of channels in that cell could no longer support a good grade of service, that cell would be sub-divided into a number of smaller cells with even lower transmitter power-fitting within the area of the former cell. The reuse pattern of channels could be repeated on a new, smaller scale, illustrated by the smaller hexagons in fig. 4.10.

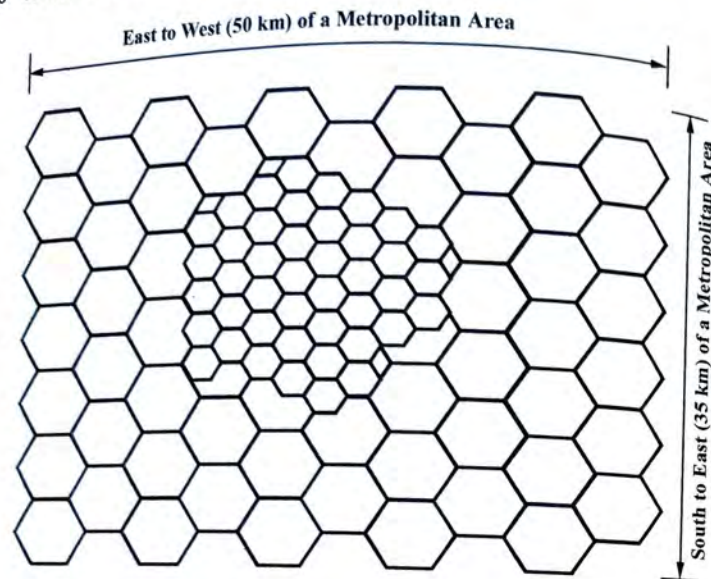


Fig. 4.10 Improvement of Capacity in a Cellular System by Cell Spacing

Another fundamental cellular idea led to the possibility of handoff control. In a cellular system, not all mobile calls may be completed within the boundaries of a single, relatively small cell. To deal with this, the cellular system is equipped with its own system level switching and control capability.

Through continuous monitoring of signal strength or of other digital parameters received from individual cell sites, the cellular system can sense when a mobile unit with a call in progress passes from one cell to another and can switch the call to the new call without interruption.

Q.26. Explain the concept of "FREQUENCY REUSE" as applied to Cellular Communications. What are the advantages of this approach? What are the consequences of choosing – A Large Cell Reuse Pattern (Cluster Size) and Small Cell Reuse Pattern (Cluster) on the following parameters –

- (i) Spectrum utilization
 - (ii) Carrier to interference ratio
 - (iii) No. of calls per cell
 - (iv) No. of channels available per cell.
- (R.G.P.V., June 2009, Dec. 2009)

Ans. Frequency Reuse – Refer the ans. of Q.25.

For the cellular mobile radio system, frequency reuse is the core concept. Users in various geographic locations (different cells) may simultaneously use the same frequency channel for this frequency reuse system. Frequency reuse system can drastically enhance the spectrum efficiency, but severe interference may take place, if the system is not properly designed. Interference because of the common use of the same channel is known as **co-channel**

interference. Frequency reuse concept is employed in the space domain and the time domain. In the time domain, frequency reuse gives in the occupation multiplexing (TDM). In the space domain, frequency reuse is classified into two types –

- (i) In two different geographic areas, same frequency allocated, such as AM or FM radio stations using the same frequency in different cities.
- (ii) In one system, same frequency repeatedly used in a same general area – the method can be employed in cellular systems. There are several co-channel cells in the system. Total frequency spectrum allocation can be classified into N frequency reuse patterns. This is shown in fig. 4.11 for $N = 4, 7, 12$, and 19 .

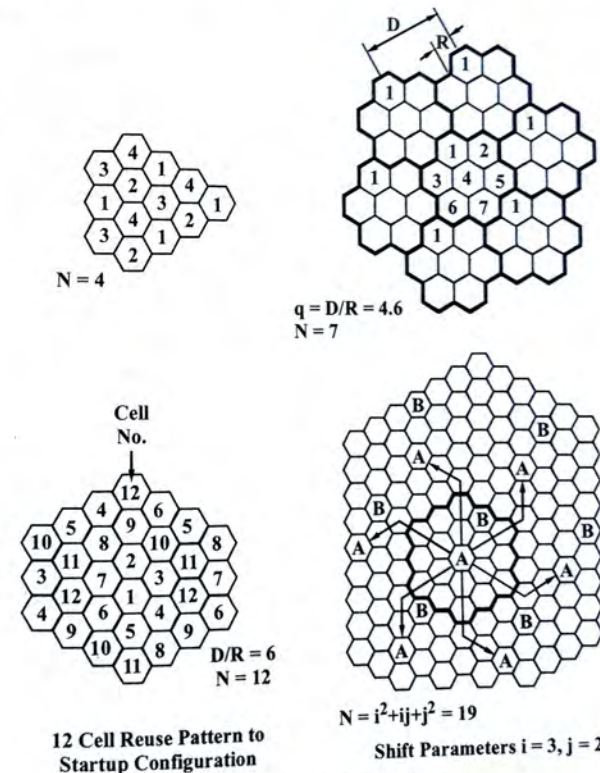


Fig. 4.11 N-cell Reuse Pattern

Minimum distance that permits the same frequency to be reused will be determined by several factors, like the type of geographic terrain contour, the number of co-channel cells in the vicinity of the centre cell, the transmitted power at every cell site and the antenna height.

Frequency reuse distance D is given as –

$$D = \sqrt{3N} R$$

where, N = Frequency reuse patterns.

$$\text{Then, } D = \begin{cases} 7.55 R, & N = 19 \\ 6 R, & N = 12 \\ 4.6 R, & N = 7 \\ 3.46 R, & N = 4 \end{cases}$$

The value of frequency reuse pattern N increases and the frequency reuse distance D increases, when all the cell sites transmit the same power. This increased D minimize the opportunity that co-channel interference may take place.

In theory, a large value of frequency reuse pattern N is required. The total number of assigned channels is constant. The number of channels allocated to each of N cells becomes small, if N is too large. Trunking inefficiency results, when total number of channels in N cells is divided as N rises. The same principle applies to spectrum inefficiency – spectrum inefficiency increases, when the total number of channels are distributed into two network systems working in the same area.

After this, the smallest number N , which can still meet our system performance needs, is obtained. This includes estimating co-channel interference and the minimum frequency reuse distance D is chosen to minimize co-channel interference. By choosing $i = 1, j = 1$ in equation $N = i^2 + ij + j^2$, the smallest value of $N = 3$, is obtained.

While designing a system, the traffic conditions in the area during a busy hour are some of the parameters which will help to determine both, the sizes of different cells and the number of channels in them.

Traffic conditions at each particular cell determine the maximum number of calls per hour per cell. The maximum number of calls per hour can be taken care of each cell after the maximum number of frequency channels per cell has been implemented in each cell. Take Q_i the maximum number of calls per hour in each cell, and sum them. It is assumed that during the busy hour, 60% of the car phones will be used, on average one call per phone ($\eta_c = 0.6$) if that phone is used. Then, the total permitted subscriber traffic M_t can be found.

The formula for determining the number of frequency reuse cells in a standard cellular configuration is derive by using the following equations

$$q = \left(6 \frac{C}{I}\right)^{\frac{1}{\gamma}} \quad \dots(i)$$

and

$$q = \sqrt{3N} \quad \dots(ii)$$

with $\gamma = 4$ based on 40 dB/decade path loss.

$$\frac{C}{I} = \frac{1}{6} \left(\frac{D}{R}\right)^4 = \frac{(3N)^2}{6} = \frac{3N^2}{2} \quad \dots(iii)$$

or

$$N = \sqrt{\frac{2C}{3I}} \quad \dots(iv)$$

The number of frequency reuse cells is a function of the required carrier-to-interference ratio. Higher carrier-to-interference ratio at the boundary of a cell result in the need for more frequency reuse cells. Thus pattern of reuse cells can be determined.

Number of Channels Available Per Cell – Another factor to be determined is the number of channels per cell, which is a function of the total number of channels available (amount of available spectrum divided by channel bandwidth) and the required carrier-to-interference ratio. The formula for this factor is

$$m = \frac{B_t}{B_c K} = \frac{B_t}{B_c \sqrt{\left(\frac{2}{3}\right) \left(\frac{C}{I}\right)}}$$

For $M = mk$ total number of channels

where m = Number of channels per cell, it is also called radio capacity.

k = Number of frequency reuse cells

B_t = Total bandwidth

B_c = Channel bandwidth.

Q.27. Explain the cellular design techniques that are needed to improve coverage and capacity in cellular systems.
(R.G.P.V., May/June 2006, Dec. 2013)

Ans. The concept of cellular telephone systems, continues by considering typical design techniques involved, illustrates the dilemma of interference and system capacity (number of users accommodated per unit area) in a cellular systems, and then examines a common technique employed to increase capacity for a fixed spectrum allocation known as trunking. Performance analysis to determine the quality of service provided for a designed trunked cellular systems is then presented. Other capacity and coverage enhancing techniques such as cell splitting, sectoring, etc., are then discussed.

The following approaches are used in cellular system to increase the coverage –

(i) **Increasing the Transmitted Power** – Usually, increasing the transmitted power of each channel results in coverage of a larger area. When the power level is doubled, the gain increases by 3 dB.

(ii) **Increasing Cell-site Antenna Height** – The 6 dB/oct rule applies to the cell-site antenna height in a flat terrain, that is, doubling the antenna height causes a gain increase of 6 dB.

(iii) **Lowering the Threshold Level of a Received Signal** – When the threshold level is lowered, the acceptable received power is lower and the radius of the cell increases.

(iv) **Diversity Receiver** – A diversity receiver is very useful in reducing the multipath fading. When the fading reduces, the reception level can be increased.

(v) By using repeaters and enhancers to enlarge the coverage area or to fill in holes.

The following approaches are used in cellular system to increase the capacity –

(i) **Small Cell Size** – If we can control the radiation pattern we can reduce the size of the cell and increase the traffic capacity. This approach is based on the assumption that all the mobile units are identical, including the mobile antennas and their mountings.

(ii) **Queuing** – Queuing of handoff calls can increase traffic capacity.

(iii) **Enhanced Frequency Spectrum** – Cellular mobile industries have been allocated an additional 166 voice channels. With an enhanced frequency spectrum, traffic capacity is increased.

(iv) **Dynamic Channel Assignment** – Dynamic, rather than fixed, channel assignment is another means of increasing traffic capacity.

Q.28. Explain the advantages and disadvantages of digital wireless system over analog wireless system.
(R.G.P.V., Dec. 2008)

Ans. Advantages – Various advantages of digital techniques in cellular systems are summarized as follows –

(i) **Advances in Digital Modulation Techniques** – The spectral utilization can be improved beyond that achieved with analog techniques by the use of new generations of digital modulation techniques.

(ii) **Robust Source and Channel Coding Techniques** – These techniques which are available for digital voice or data transmission, enhance performance in the mobile environment.

(iii) **Lower Bit-rate Digital Voice Coding** – The latest improvements in low bit-rate voice coding, together with digital modulation techniques, enable the inclusion of several voice channels onto one carrier and therefore improving spectral utilization.

(iv) **More Robust to Interference** – Digital systems have an improved performance in high co-channels interference (CCI) and adjacent channel interference (ACI) environments as compared with analog systems. Digital systems are likely to operate under conditions of much higher co-channel interference which will enable designers to bring the cell size and the reuse distance down and even to reduce the reuse pattern.

(v) **Reduction of Overhead for Signaling** – In the 333 channels originally allocated for AMPS, about 21 were required for call setup. The analog systems are inefficient in this regard. This overhead reduces the usable capacity of a bandwidth limited system. The overhead could be greatly reduced by using digital techniques for synchronization, control messages and performance monitoring.

(vi) **Inclusion of New Services** – The digital technology makes it possible to introduce new services that are not supported by AMPS or other analog systems.

(vii) **Flexible Bandwidth** – A predetermined fixed radio frequency bandwidth could lead to inefficient use of the spectrum by not permitting users to adjust their bandwidth and timing to meet their actual communications requirements.

(viii) **Improved Efficiency of Access and Handoff Control** – For a fixed allocation of spectrum, a large increase in capacity implies corresponding reductions in cell size. Its means that signalling activity increases as more rapid handoff occur. To handle more access requests and registrations from the community of portables and mobiles in each cell, the base stations are required.

Disadvantages – The problem of a rapidly expanding customer base while at the same time the allocated frequency spectrum remained unchanged faced by telephone companies. As is evident with N-AMPS, user capacity can be expanded by subdividing existing channels (band splitting) partitioning cells into smaller subcells (cell splitting), and modifying antenna radiation patterns (sectoring). However, the degree of subdivision and redirection is limited by the complexity and amount of overhead required to process handoffs between cells. Another serious restriction is the availability and cost of purchasing or leasing property for cell sites in the higher density traffic areas. The transmission of digitally encoded analog signals requires significantly more bandwidth than simply transmitting the original analog signals. Bandwidth is one of the most

important aspects of any communications system because it is costly and limited. Analog signals must be converted to digital pulses prior to transmission and converted back to their original analog form at the receiver, thus necessitating additional encoding and decoding circuitry. Digital transmission requires precise time synchronization between the clocks in the transmitters and receivers. Finally, digital transmission systems are incompatible with older analog transmission systems.

Q.29. What are the advantages and disadvantages of having different size cells for wireless networking ?
(R.G.P.V., Dec. 2005)

Ans. Advantages –

(i) **Higher Capacity** – Implementing SDM permits frequency reuse. When one transmitter is far away from another, i.e., outside the interference range, it can reuse the same frequencies. Since most mobile phone systems allocate frequencies to certain users or certain hopping patterns, this frequency is blocked for other users. But, frequencies are a scarce resource, and the number of concurrent users per cell is very limited. Huge cells do not permit for more users. On the other hand, they are limited to less possible users per km. This is also the reason for using very small cells in cities where many more people use mobile phones.

(ii) **Less Transmission Power** – However, power aspects are not a major problem for base stations, but they are indeed problematic for mobile stations. A receiver far away from a base station would require much more transmit power than the current few watts. But, for mobile handheld devices, energy is a serious problem.

(iii) **Local Interference Only** – When sender and receiver are at long distance, then it results in interference problems. By using small cells, mobile stations and base stations only have to deal with 'local' interference.

(iv) **Robustness** – Cellular systems are decentralized, and therefore more robust against the failure of single components. In case one antenna fails, it only affects communication within a small area.

Disadvantages –

(i) **Infrastructure Required** – Wireless networks require a complex infrastructure to connect all base stations. This includes several antennas, switches for call forwarding, location registers to find a mobile station, etc., that makes the entire system quite costly.

(ii) **Handover Required** – Mobile station has to perform a handover when changing from one cell to another. Depending on the cell size and the speed of movement, this can occur quite often.

(iii) **Frequency Planning** – Frequencies have to be distributed carefully to avoid interference between transmitters. On the one hand interference should be avoided, on the other hand only a limited number of frequencies is available.

Q.30. How a call initiated by a mobile is established and how a call to a mobile user initiated by a landline subscriber is established in cellular mobile system.

Ans. Mobile-to-mobile Call Procedures –

(i) The originating mobile unit initiates the call in the same manner as it would for a mobile to wireline call.

(ii) The cell-site controller receives the caller's identification number and the destination telephone number through a reverse control channel, which are then forwarded to the MTSO.

(iii) The MTSO sends a page command to all cell-site controllers to locate the destination party.

(iv) Once the destination mobile unit is located, the destination cell-site controller sends a page request through a control channel to the destination party to determine if the unit is on or off hook.

(v) After receiving a positive response to the page, idle user channels are assigned to both mobile units.

(vi) Call progress tones are applied in both directions.

(vii) When the system receives notice that the called party has answered the telephone, the switch terminates the call progress tones, and the conversation begins.

(viii) If a mobile subscriber wishes to initiate a call and all user channels are busy, the switch sends a directed retry command, instructing the subscribers unit to reattempt the call through a neighbouring cell.

(ix) If the system cannot allocate user channels through a neighbouring cell, the switch transmits an intercept message to the calling mobile unit over the control channel.

(x) If the called party is off hook, the calling party receives a busy signal.

(xi) If the called number is invalid, the calling party receives a recorded message announcing that the call cannot be processed.

Wireline-to-mobile Call Procedures –

(i) The wireline telephone goes off hook to complete the loop, receives a dial tone, and then inputs the mobile unit's telephone number.

(ii) The telephone number is transferred from the PSTN switch to the cellular network switch (MTSO) that services the destination mobile number.

(iii) The cellular network MTSO receives the incoming call from the PSTN, translates the received digits, and locates the base station nearest the mobile unit, which determines if the mobile unit is on or off hook.

(iv) If the mobile unit is available, a positive page response is sent over a reverse control channel to the cell site controller, which is forwarded to the network switch.

(v) The cell-site controller assigns an idle user channel to the mobile unit and then instructs the mobile unit to tune the selected channel.

(vi) The mobile unit sends, verification of channel tuning through the cell site controller.

(vii) The cell-site controller sends an audible call progress tone to the subscribers mobile telephone, causing it to ring. At the same time, a ring back signal is sent back to the wireline calling party.

(viii) The mobile answers the switch terminates the call progress tones, and the conversation begins.

Q.31. Define the following terms of cellular mobile system –

(i) Cell area

(ii) Reuse distance

(iii) Reuse frequency

(iv) Co-channel and adjacent channel interference.

(R.G.P.V., Dec. 2006)

Ans. (i) Cell Area – The FCC originally defined geographic cellular radio coverage areas on the basis of modified 1980 census figure. With the cellular concept, each area is further divided into hexagonal-shaped cells that fit together to form a honeycomb pattern. The hexagon shape was chosen because it provides the most effective transmission by approximating a circular pattern while eliminating gaps inherently present between adjacent circles. Each geographical area is allocated a fixed number of cellular voice channels.

(ii) Reuse Distance – The minimum distance that permits the same frequency to be reused will be determined by several factors, like the number of co-channel cells in the vicinity of the centre cell, the type of geographic terrain contour, the transmitted power at each cell site and the antenna height.

(iii) Reuse Frequency – When an allocated frequency band is filled, additional capacity can be achieved by reuse of the frequency spectrum. By increasing the size of an antenna, beam width of the antenna is also reduced. Thus, different beams of the same frequency can be directed to different geographical areas of earth. This is called frequency reuse.

(iv) Co-channel and Adjacent Channel Interference – When frequency reuse is implemented, several cells with in a given coverage area use the same set of frequencies. Two cells using the same set of frequencies are called co-channel cells and the interference between them is called co-channel interference. Unlike thermal noise co-channel interference cannot be reduced by simply increasing transmit powers because increasing the transmit power in one cell increase the likelihood of that cell's transmission interfering with another cell's transmission. To reduce co-channel interference a certain minimum distance must separate co-channels.

Adjacent channel interference occurs when transmission from adjacent channels interfere with each other. Adjacent channel interference results from imperfect filters in receivers that allow nearby frequencies to enter the receiver. Adjacent channel interference is most prevalent when an adjacent channel is transmitting very close to a mobile units receiver at the same time the mobile unit is trying to receive transmissions from the base station on the adjacent frequency. This is called the near far effect and is most prevalent when a mobile unit is receiving a weak signal from the base station.

Q.32. Discuss the various effects of mobile radio propagation.

(R.G.P.V., Dec. 2013, 2016)

Ans. Mobile radio communication introduces complexities not found in wire communication or in fixed wireless communication. Two general areas of concern are signal strength and signal propagation effects.

(i) Signal Strength – The strength of the signal between the base station and the mobile unit must be strong enough to maintain signal quality at the receiver but not so strong as to create too much co-channel interference with channels in another cell using the same frequency band. Several complicating factors exist. Human-made noise varies considerably, resulting in a variable noise level. For example automobile ignition noise in the cellular frequency range is greater in the city than in a suburban area. Other signal sources vary from place-to-place. The signal strength varies as a function of distance from the BS to a point within its cell. Moreover, the signal strength varies dynamically as the mobile unit moves.

(ii) Fading – Even if signal strength is within an effective range, signal propagation effects may disrupt the signal and cause errors. Usually at every $\lambda/2$ space, interval fade may occur.

Q.33. Define types of radio propagation mechanisms. (R.G.P.V., Dec. 2015)

Ans. Radio signals with frequencies above 800 MHz, used in the wireless networks, have extremely small wavelengths compared with the dimensions of building features, so electromagnetic waves can be treated simply as rays.

This means that ray-optical methods can be used to describe the propagation within and even outside buildings by treating electromagnetic waves as traveling along localized ray paths. The fields associated with the ray paths change sequentially based on the features of the medium that the ray encounters.

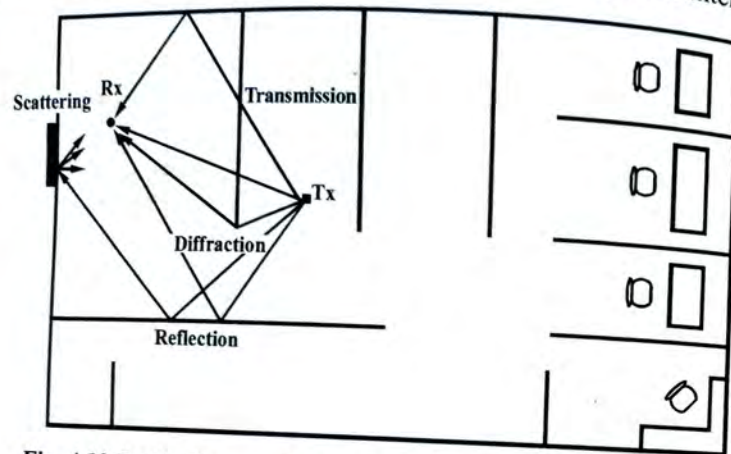


Fig. 4.12 Radio Propagation Mechanisms in an Indoor Area

In order to describe radio propagation with ray optics, three basic mechanisms are generally considered while ignoring other complex mechanisms. These mechanisms are illustrated in figs. 4.12 and 4.13 for indoor and outdoor applications, respectively.

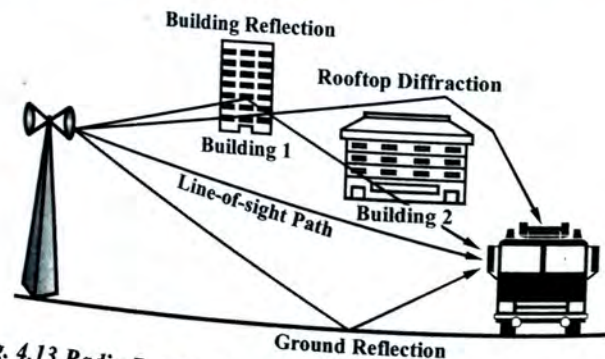


Fig. 4.13 Radio Propagation Mechanisms in an Outdoor Area

(i) **Reflection and Transmission** – Specular reflections and transmission occur when electromagnetic waves impinge on obstructions larger than the wavelength. Usually rays incident upon the ground, walls of buildings, the ceiling, and the floor undergo specular reflection and transmission with the amplitude coefficients usually determined by plane wave analysis. Upon reflection or transmission, a ray attenuates by factors that depend on the

frequency, the angle of incidence, and the nature of the medium (its material properties, thickness, homogeneity, etc.). These mechanisms often dominate radio propagation in indoor applications. In outdoor urban areas, this mechanism often loses its importance because it involves multiple transmissions that reduce the strength of the signal to negligible values.

(ii) **Diffraction** – Rays that are incident upon the edges on buildings, walls, and other large objects can be viewed as exciting the edges to act as a secondary line source. Diffracted fields are generated by this secondary wave source and propagate away from the diffracting edge as cylindrical waves. In effect, this results in propagation into *shadowed* regions because the diffracted field can reach a receiver, which is not in the line of sight of the transmitter. Because a secondary source is created, it suffers a loss much greater than that experienced via reflection or transmission. Consequently, diffraction is an important phenomenon outdoors where signal transmission through buildings is virtually impossible. It is less consequential indoors where a diffracted signal is extremely weak compared to a reflected signal or a signal that is transmitted through a relatively thin wall.

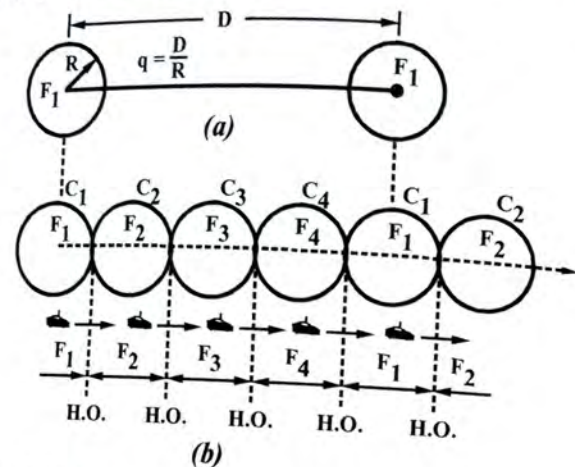
(iii) **Scattering** – Irregular objects such as walls with rough surfaces and furniture (indoors) and vehicles, foliage, and the like (outdoors) scatter rays in all directions in the form of spherical waves. This particularly occurs when objects are of dimensions that are on the order of a wavelength or less of the electromagnetic wave. Propagation in many directions results in reduced power levels, especially far from the scatterer. As a result, this phenomenon is not that significant unless the receiver or transmitter is located in a highly cluttered environment. This mechanism dominates diffused IR propagations when the wavelength of the signal is such that the roughness of the wall results in extensive scattering. In satellite and mobile radio applications, foliage often causes scattering.

Q.34. What is the process of handoff? Define types of handoff.
(R.G.P.V., Dec. 2015)

Ans. For understand the handoff concept, it is straightforward to employ a one-dimensional illustration as shown in fig. 4.14, however a real two-dimensional cellular configuration would cover an area with cells.

In fig. 4.14 (a), two co-channel cells using the frequency F_1 separated by a distance D are shown. The distance D and the radius R can be controlled with the value of q . Now, in order to give a communication system in the entire area, we have to fill in with other frequency channels, like F_2 , F_3 and F_4 . As shown in fig. 4.14 (b), the fill-in frequencies between two co-channel cells.

F_2 , F_3 and F_4 can also be allocated to their corresponding cells C_2 , C_3 and C_4 in proportion to the same value of q .



(a) Co-channel Interference Reduction Ratio (q) (b) Fill-in Frequencies
Fig. 4.14 Handoff Phenomenon (Mechanism)

If a mobile unit is starting a call in cell C_1 and then moves to cell C_2 . While the mobile unit moves from cell C_1 to cell C_2 , the call is dropped and reinitiated in the frequency channel from F_1 to F_2 . This process of converting frequencies is performed without the user's intervention. This process of handoff is continued in the cellular system. For any successful mobile system the handoff processing method is an important task.

Handoff can be required in two conditions where the cell site receives weak signals from the mobile unit –

- At the cell boundary, say, -100 dBm which is level for requesting a handoff in a noise limited environment.
- When the mobile unit is reaching the signal power holes within the cell site.

There are two types of handoff –

(i) **Soft Handoff** – The MS commences communication with a new base station without interrupting communication with the old base station. Soft handoff means an identical frequency assignment between the old BS and new base station. Soft handoff provides different-site selection diversity to enhance the signal.

(ii) **Hard Handoff** – Hard handoff means a different frequency assignments between the old BS and new BS. The MS transmits between two base station with different frequency assignments.

Q.35. Explain handoff techniques in cellular networks.

(R.G.P.V., June 2004, 2008, Dec. 2009)

Or

What is handoff? Explain different handoff initialization techniques with suitable diagram.

(R.G.P.V., Dec. 2016)

Or

Explain handoff.

(R.G.P.V., Dec. 2012)

Ans. The procedure for changing the allotment of a mobile unit from one BS to another is handoff.

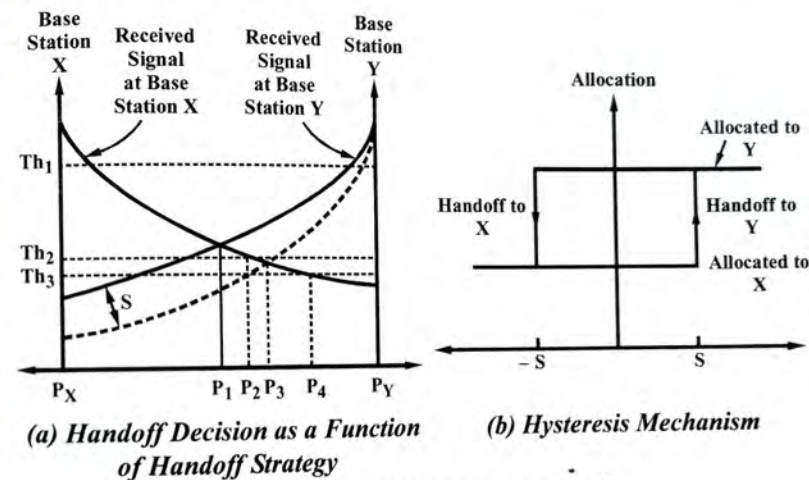


Fig. 4.15 Handoff between Two Cells

Fig. 4.15 explains various handoff techniques –

(i) **Relative Signal Strength** – If the signal strength at Y first exceeds that at X, then the mobile unit is handed off from BS X to BS Y. The mobile unit is handed back to X when the signal strength at Y subsequently falls below that of X. In fig. 4.15 (a), handoff happens at point L_1 . At this point, signal strength to BS X is declining but still enough. Even with power averaging, this strategy can lead to a ping-pong effect in which the unit is repeatedly passed back and forth between two BSs as signal strength fluctuates due to multipath effects.

(ii) **Relative Signal Strength with Threshold** – Handoff only happens when (a) the other signal is the stronger of the two and (b) the signal at the current BS is sufficiently poor. The intention is that handoff is unnecessary so long as the signal at the current BS is adequate. This scheme performs similar to the relative signal strength scheme when a high threshold is used like Th_1 . Handoff happens at P_2 with a threshold of Th_2 . The mobile unit may

move far into the new call (P_4) before handoff when the threshold is set quite low than the crossover signal strength like Th_3 . This may result in a dropped call and reduces the quality of the communication link. A threshold effectiveness relies on prior knowledge of the crossover signal strength between the current and candidate base stations, so a threshold should not be utilized alone.

(iii) **Relative Signal Strength with Hysteresis** – Only when the new base station is sufficiently stronger compared to the current one then handoff happens. Handoff happens at P_3 in this case. This approach stops the ping-pong effect, as once handoff happens, the effect of the margin S is reversed. The term hysteresis means a phenomenon called as relay hysteresis and can be appreciated with the help of fig. 4.15 (b). The handoff mechanism can be thought of as having two states. The mechanism will produce a handoff when the relative signal strength approaches or exceeds while the mobile unit is allocated to BS X. Once the mobile unit is allocated to Y, it remains so until the relative signal strength drops down – 8, at which point it is given back to X. The drawback of this strategy is that the first handoff may still be unnecessary if BS X still has enough signal strength.

(iv) **Relative Signal Strength with Hysteresis and Threshold** – Handoff happens only if (a) the target base station is stronger compared to the current one by a hysteresis margin, and (b) the current signal level falls below a threshold. In this example, handoff happens at P_4 if the threshold is at Th_3 and at P_3 if the threshold is either Th_1 or Th_2 .

(v) **Prediction Techniques** – The handoff decision depends on the expected future value of the received signal strength.

Q.36. How the power is controlled in a cellular system? Explain the difference between open-loop and closed-loop power control.

Ans. In a cellular system, number of design issues make it desirable to include a dynamic power control capability –

(i) The received power must be sufficiently above the background noise for effective communication, which dictates the required transmitted power. As the mobile unit moves away from the transmitter, the received power declines due to normal attenuation. In addition, the effects of reflection, diffraction, and scattering can cause rapid changes in received power levels over small distances. This is because the power level is the sum from signals coming from a number of different paths and the phases of those paths are random, sometimes adding and sometimes subtracting. As the mobile unit moves, the contribution along various paths change.

(ii) At the same time, it is desirable to minimize the power in the transmitted signal from the mobile unit, to reduce co-channel interference, alleviate health concerns and save battery power.

(iii) In spread spectrum (SS) systems using code division multiple access (CDMA), it is desirable to equalize the received power level from all mobile units at the BS. This is crucial to system performance because all users have the same frequency allocation.

Cellular system use the two kinds of power control.

(i) **Open-loop Power Control** – It depends solely on the mobile unit, with no feedback from the BS, and is used in some SS systems. In SS systems, the BS continuously transmits an unmodulated signal, known as a pilot. The pilot allows a mobile unit to acquire the timing of the forward (BS to mobile) CDMA channel and provides a phase reference for demodulation. It can also be used for power control. The mobile unit monitors the received power level of the pilot and sets the transmitted power in the reverse (mobile to BS) channel inversely proportional to it. This approach assumes that the forward and reverse link signal strengths are closely correlated, which is generally the case. The open-loop approach is not as accurate as the closed-loop approach. However, the open-loop scheme can react more quickly to rapid fluctuations in signal strength, such as when a mobile unit emerges from behind a large building. This fast action is required in the reverse link of a CDMA system. Where the sudden increase in received strength at the BS may suppress all other signals.

(ii) **Closed-loop Power Control** – It adjusts signal strength in the reverse (mobile to BS) channel based on some metric of performance in that reverse channel, such as received signal power level, received signal-to-noise ratio, or received bit error rate. The BS makes the power adjustment decision and communicates a power adjustment command to the mobile unit on a control channel. Closed-loop power control is also used to adjust power in the forward channel. In this case, the mobile unit provides information about received signal quality to the BS, which then adjusts transmitted power.

Q.37. What is sectorization?

Ans. The total number of available channels can be divided into subgroups depending on the sectorization of the cell configuration, the 120° sector groups depending on the sectorization of the cell configuration, the 120° sector system, the 60° sector system and the 45° sector system. Usually a 7 cell system uses three 120° sectors per cell, with the total number of channel subgroups being 21. The sector angle can be narrowed in certain locations and special conditions in order to assign more channels in one sector without increasing neighbouring channel interference. The sectorization is useful for the same purpose as the channel borrowing technique in delaying cell splitting. In sectorization, channel coordination to avoid co-channel interference is much easier than in cell splitting. The trunking efficiency decreases in sectorization for the given same number of channels.

Q.38. Give the comparison of omni-cells and sectorized cells.

Ans. The comparison of omni-cells (non-sectorized cells) and sectorized cells are given as under –

(i) **Omni-cells (Non-sectorized cells)** – When seven cell frequency reuse pattern is used, the frequency subgroups are assigned in each cell can be followed by the frequency management chart. However, terrain is seldom flat; thus twelve cell frequency reuse pattern is sometimes required for reducing co-channel interference. For $N = 12$ the channel reuse distance is $D = 6R$ or the co-channel reduction factor $q = 6$.

(ii) **Sectorized Cells** – These are categorized into three types –

(a) For both transmitting and receiving sectorization, the 120° sector cell is used.

(b) The 60° sector cell is used for both transmitting and receiving sectorization. Changing sectors during a cell needs handoffs. Greater handoffs are expected for a 60° sector than a 120° sector in areas close to cell sites.

(c) For receiving sectorization only the 120° sector or 60° sector cell is used. In this case the transmitting antenna is omni-directional. In this cell the number of channels is not subdivided for each sector. Thus no handoffs are required when changing sectors. This receiving-sectorization only configuration does not decrease interference or increase the D/R ratio; it only allows for a more accurate conclusion regarding handing off the calls to neighbouring cells.

Q.39. Write down the difference between cellular network and Wi-Fi networks.

Ans. Difference between cellular and Wi-Fi networks –

S.No.	Wi-Fi Network	Cellular Network
(i)	Wi-Fi is a local area networking technology that uses radio waves to provide high-speed Internet access to mobile devices enabled with Wi-Fi.	Cellular plan are attached to the device that uses cellular signal to connect to the Internet.
(ii)	Wi-Fi the name given to the wireless devices that are certified to use the 802.11 standard for communication.	The main objective of a cellular communication network is to provide high-capacity coverage over a wide area.
(iii)	Wi-Fi enables devices use Wi-Fi to send and receive wireless signal and cannot accept SIM.	Cellular network is offered by mobile network operators or service providers.

(iv)	Wi-Fi spans a limited area within a specified range.	Cellular network spans a wide area covering cities, towns, etc.
(v)	Wi-Fi is typically much faster than a cellular network.	Cellular network is relatively slower than Wi-Fi in terms of speed and reliability.
(vi)	Wi-Fi has no limit on amount of data and is typically based on the Wi-Fi network.	Cellular data is based on monthly data plans that come with limited amount of data.

GSM – SERVICES, SYSTEM ARCHITECTURE, LOCALIZATION AND CALLING, HANDOVER AND ROAMING

Q.40. Explain GSM.

(R.G.P.V., Dec. 2012)

Ans. **GSM** is the most successful digital mobile telecommunication system in the world today. It is used by over 800 million people in more than 190 countries. In 1980s, Europe had numerous coexisting analog mobile phone systems, which were often based on similar standards, but ran on slightly different carrier frequencies. To avoid this situation for a second generation fully digital system, the group special mobile (GSM) was founded in 1982. Soon, this system given named the *global system for mobile communication (GSM)* with the specification process lying in the hands of ETSI (European Telecommunication Standard Institute), a GSM association 2002. In the context of UMTS (Universal mobile telecommunication system) and the creation of 3GPP (Third generation partnership project) the whole development process of GSM was transferred to 3GPP and further development is combined with 3G development. 3GPP assigned new numbers to all GSM standards.

The primary goal of GSM was to provide a mobile phone system that allows users to roam and provides voice services compatible to ISDN and other PSTN systems.

GSM is a typical second generation system, replacing the first generation analog systems, but not offering the high worldwide data rates that the third generation systems, such as UMTS, are promising. GSM has initially been deployed in Europe using 890-915 MHz for uplinks and 935-960 MHz for downlinks this system is now also called GSM 900 to distinguish it from the later versions. These versions comprise GSM at 1800 MHz (1710-1785 MHz uplink, 1805-1880 MHz downlink), also called DCS (digital cellular system) 1800, and the GSM system mainly used in the US at 1900 MHz (1850-1910 MHz uplink, 1930-1990 MHz downlink), also called PCS (personal communications service) 1900. Two more versions of GSM exist. GSM 400 is

a proposal to deploy GSM at 450.4-457.6/478.8-486 MHz for uplinks and 460.4-467.6/488.8-496 MHz for downlinks. This system could replace analog systems in sparsely populated areas.

Q.41. List the important GSM air-interface standard specifications.
(R.G.P.V., Dec. 2015)

Ans. The important GSM air-interface standard specifications are as follows –

S.No.	Parameter	Specifications
(i)	Frequency band	Uplink 890-915 MHz Downlink 935-960 MHz
(ii)	Spectral allocation	50 MHz (25 MHz for uplink and 25 MHz for downlink)
(iii)	RF channel bandwidth	200 kHz or 0.2 MHz
(iv)	Multiple access technique	FDMA and TDMA
(v)	Forward and reversed channel frequency spacing	45 MHz
(vi)	Duplexing technique	Frequency division duplexing (FDD)
(vii)	Transmission and receiving time slot spacing	3 times slots
(viii)	ARFCN number	0-124 and 975-1023
(ix)	Number of voice channels	1000
(x)	Time slot period	577 μ s
(xi)	Modulation scheme	GMSK
(xii)	Modulation data rate	270.833 kbps
(xiii)	Bit period	3.692 μ s
(xiv)	Frame period	4.615 ms
(xv)	Number of time slots per RF channel bandwidth	8
(xvi)	Type of equalizers	Adaptive equalizer
(xvii)	Interleaving (max-delay)	40 ms
(xviii)	Handheld mobile transmission power	1 W_{max} ; 125 m W_{avg} .
(xix)	Channel coding	CRC with $r = 1/2$; L = 5 convolution coding
(xx)	Speech coding	REL-P-LTP @ 13.4 kbps
(xxi)	Spectrum efficiency	1.35 bps/Hz

Q.42. Discuss the different types of mobile services provided by GSM.
Explain each in brief.
(R.G.P.V., Dec. 2013, 2016)

Or

What are the types of GSM services ? Explain. (R.G.P.V., Dec. 2014)

Or

What are the various mobile services provided by GSM ? Explain each in brief.
(R.G.P.V., Nov. 2018)

Ans. GSM allows the integration of different voice and data services and the interworking with existing networks. GSM had defined three different categories of services—bearer, tele and supplementary. A reference model for GSM services is shown in fig. 4.16.

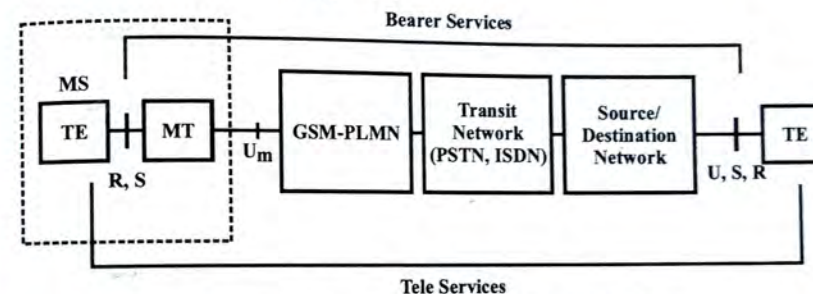


Fig. 4.16 Reference Model for GSM Services

A mobile station (MS) is connected to the GSM public land mobile network (PLMN) via the U_m interface. This network is connected to transit networks, e.g., integrated services digital network (ISDN) or traditional public switched telephone network (PSTN). There must be an additional network, the source/destination network, before another terminal TE is connected.

Now bearer services comprise all services which enable the transparent transmission of data between the interfaces to the network, that is, S in case of the mobile station, and a similar interface for the other terminal. In case of ISDN, U, S and R interfaces have not been defined for all networks. Therefore, it depends on the specific network that the interface is used as a reference for the transparent transmission of data.

Tele services are application specific and may thus need all seven layers of the ISO/OSI reference model. These services are specified end-to-end, that is, from one terminal TE to another.

(i) Bearer Services – Different mechanisms are specified by GSM for data transmission. Original GSM permits for data rates of upto 9600 bit/sec for non-voice services. Bearer services allow transparent and non-transparent, synchronous or asynchronous data transmission.

Transparent bearer services only use the functions of the physical layer (layer 1) to send data. Therefore, data transmission has a constant delay and throughput if no transmission errors take place. The only mechanism to enhance transmission quality is the use of forward error correction (FEC), which codes redundancy into the data stream and thus helps to reconstruct the original data in case of transmission errors. Depending on the FEC, data rates of 2.4, 4.8, and 9.6 kbits/sec are possible.

In order to implement error correction and flow control, the non-transparent bearer services use protocols of the layers two and three. These services use the transparent bearer services, adding a radio link protocol (RLP). This protocol has mechanisms of high-level data link control (HDLC) and special selective-reject mechanisms to trigger retransmission of erroneous data. The obtained bit error rate is less than 10^{-7} , but now throughput and delay may vary depending on transmission quality.

With the help of transparent and non-transparent services, GSM specifies many bearer services for interworking with PSTN, ISDN, and packet switched public data network (PSPDN). Data transmission can be full-duplex, synchronous with data rates of 1.2, 2.4, 4.8, and 9.6 kbits/sec or full-duplex, asynchronous from 300 to 9,600-bits/sec. Obviously, these relatively low data rates reflect the assumption that data services will only constitute some small percentage of the overall traffic.

(ii) **Tele Services** – The main focus of GSM is on voice-oriented tele services. These contain encrypted voice transmission message services and basic data communication with terminals as known from the PSTN or ISDN. But, the main service is telephony, the primary goal of GSM was the provision of high-quality digital voice transmission, offering at least the typical bandwidth of 3.1 kHz of analog phone systems. Special codecs are used for voice transmission, while other codecs are used for transmission of analog data for communication with traditional computer modems.

Emergency number is the another service offered by GSM. This service is mandatory for all providers and free of charge. This connection also has the highest priority, possibly pre-empting other connections. This will automatically be set up with the closest emergency centre.

Short message service (SMS) is the another useful service which is used for very simple message transfer. This offers transmission of messages of upto 160 characters. SMS messages do not use the standard data channels of GSM but exploit unused capacity in the signalling channels. Therefore, sending and receiving of SMS is possible during data or voice transmission.

Another worldwide non-voice tele service is group 3 fax. In this service, fax data is sent as digital data over the analog telephone network according to the ITU-T standards T.4 and T.30 using modems.

(iii) **Supplementary Services** – The GSM providers can offer supplementary services. These services provide various enhancements for standard telephony service. Services offered can vary from provider to provider. Typical services are user identifications call re-direction, or forwarding of ongoing calls.

Q.43. Name the main elements of the GSM system architecture and describe their functions.
(R.G.P.V., Dec. 2010, 2017)

Or

Draw and explain functional architecture of GSM system.

(R.G.P.V., Dec. 2011)

Or

Draw and explain the GSM network architecture. (R.G.P.V., Dec. 2015)

Or

Describe the GSM network architecture in detail with the help of diagram.
(R.G.P.V., Dec. 2016)

Ans. A simplified overview of the GSM system is shown in fig. 4.17. A GSM system has three subsystems, the radio subsystem (RSS), the network and switching subsystem (NSS), and the operation subsystem (OSS).

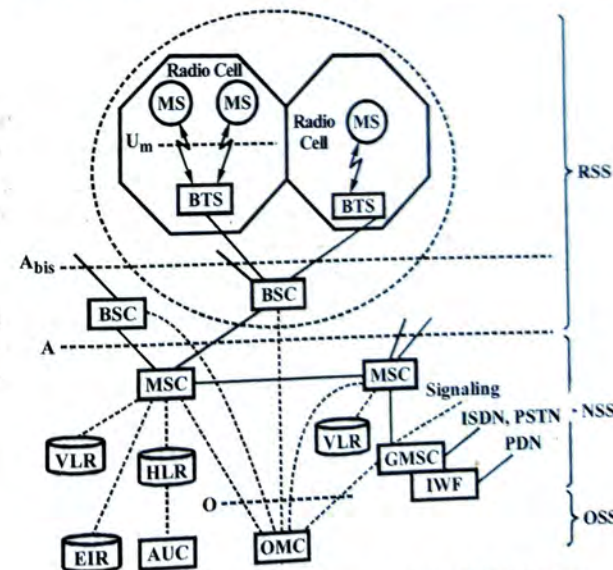


Fig. 4.17 Functional Architecture of a GSM System

(i) **Radio Subsystem** – The radio subsystem (RSS) consists of all radio specific entities, that is, the mobile stations (MS) and the base station

subsystem (BSS). Fig. 4.17 depicts the connection between the RSS and the NSS through the A interface (solid lines) and the connection to the OSS through the O interface (dashed lines). The A interface depends on circuit-switched PCM-30 systems, carrying upto 30-64 kbit/s connections, in contrast the O interface depends on X.25 carrying management data to/from the RSS.

(a) Base Station Subsystem (BSS) – A GSM network consists of many BSSs. Each BSS is controlled by a base station controller (BSC). All functions required to maintain radio connections to an MS are performed by BSS. Apart from a BSC, the BSS has many BTSs.

(b) Base Transceiver Station (BTS) – A BTS consists of all radio equipment, that is, antennas, signal processing, amplifiers required for radio transmission. A BTS can make a radio cell or many cells using sectorized antennas. A BTS is connected to BSC through the A_{bis} interface, and to the MS through the U_m interface. The A_{bis} interface comprises of 16 or 64 kbit/s connections. The U_m interface has all the mechanisms required for wireless transmission.

(c) Base Station Controller (BSC) – The BSC is used to maintain the BTSs. It performs paging of the MS, handles the handover from one BTS to another within the BSS, and reserves radio frequencies. Also, the BSC multiplexes the radio channels onto the fixed network connections at the A interface.

(d) Mobile Station (MS) – The MS consists of all user equipment and software required for communication with a GSM network. An MS comprises user independent hardware and software and of the subscriber identity module (SIM), which stores all user-specific data that is relevant to GSM.

(ii) Network and Switching Subsystem – The network and switching subsystem (NSS) is the heart of the GSM system. It connects the wireless network with standard public networks, consists of functions for worldwide localization of users and supports charging, accounting, and roaming of users between different providers in different countries, performs handovers between different BSSs. The NSS is composed of the following switches and databases –

(a) Home Location Register (HLR) – The most important database in a GSM system is the HLR. The HLR stores all user-relevant information. It consists of static information, like the *mobile subscriber ISDN number (MSISDN)*, subscribed services, and the international mobile subscriber identify (IMSI). Dynamic information is also required, such as the current *location area (LA)* of the MS, the *mobile subscriber roaming number (MSRN)*, the current VLR and MSC. The information in the HLR is updated as soon as an MS leaves its current LA. This information is required to localize a user in the worldwide GSM network.

(b) Mobile Services Switching Center (MSC) – These are high performance digital ISDN switches. They form the fixed backbone network of a GSM system, and establish connections to other MSCs and to the BSCs through the A interface. An MSC maintains a number of BSCs in a geographical region. An MSC can connect to *public data networks (PDN)* like X.25 using additional interworking functions (IWF). A *gateway MSC (GMSC)* has additional connections to other fixed networks, like PSTN and ISDN. Using standard signalling system No.7(SS7), an MSC handles all signalling required for connection setup, connection release and handover of connections to other MSCs.

(c) Visitor Location Register (VLR) – It is a dynamic database, which stores all important information required for the MS users currently in the LA that is associated to the MSC.

(iii) Operation Subsystem – The operation subsystem (OSS) consists of the required functions for network operation and maintenance. The OSS contains network entities of its own and accesses other entities through SS7 signalling. There are following entities –

(a) Authentication Centre (AuC) – A separate AuC has been defined to protect user identity and data transmission because the radio interface and mobile stations are particularly vulnerable. The AuC has the algorithms for authentication as well as the keys for encryption and produces the values required for user authentication in the HLR.

(b) Operation and Maintenance Center (OMC) – The OMC monitors and controls all other network entities through the O interface. The OMC is responsible for traffic monitoring, status reports of network entities, subscriber and security management, or accounting and billing. The concept of telecommunication management network (TMN) is used by the OMC.

(c) Equipment Identify Register (EIR) – It is a database for all IMEIs. This means that it stores all device identifications registered for this network. Because MSs are mobile, they can be easily stolen. Anyone could use the stolen MS with a valid SIM. The EIR contains a blacklist of stolen devices. The EIR also has a list of valid IMEIs, and a list of malfunctioning devices.

Q.44. Draw the GSM architecture and explain the call establishment in GSM using the logical channel. (R.G.P.V., Dec. 2014)

Ans. GSM Architecture – Refer the ans. of Q.43.

In fig. 4.18, the logical channel used for each step is also identified. Call establishment is made through the common control random access channel (RACH) and access grant channel (AGCH). Calling number and security is through the dedicated standalone dedicated control channel (SDCCH),

signalling for connection status through dedicated FACCH and traffic exchange through TCH (traffic channel).

S.No.	Steps	MS	BTS	BSC	MSC
(i)	Channel request (RACH)	→	→	→	
(ii)	Channel assigned (AGCH)	←	←	←	
(iii)	Call establishment request (SDCCH)	→	→	→	
(iv)	Authentication request (SDCCH)	←	←	←	
(v)	Authentication response (SDCCH)	→	→	→	
(vi)	Ciphering command (SDCCH)	←	←	←	
(vii)	Ciphering ready (SDCCH)	→	→	→	
(viii)	Send destination address (SDCCH)	→	→	→	
(ix)	Routing response (SDCCH)	←	←	←	
(x)	Assign traffic channel (SDCCH)	→	→	→	
(xi)	Traffic channel established (FACCH)	←	←	←	
(xii)	Available/busy signal (FACCH)	←	←	←	
(xiii)	Call accepted (FACCH)	→	→	→	
(xiv)	Connection established (FACCH)	→	→	→	
(xv)	Information exchange (TCH)	→	→	→	

Fig. 4.18 Logical Channel used for Call Establishment

Q.45. What is the function of following switches and databases in network and switching subsystem of GSM.

(i) Mobile services switching center

(ii) Home location register.

(R.G.P.V., Dec. 2012)

Ans. Refer the ans. of Q.43.

Q.46. Explain the architecture of GSM. Which types of different services does GSM offer? Give examples.

(R.G.P.V., June 2011)

Ans. Refer the ans. of Q.43 and Q.42.

Q.47. With the help of diagram, explain system architecture of GSM. Also show the interfacing between different subsystems along with components of each subsystem.

(R.G.P.V., Nov. 2019)

Ans. Refer the ans. of Q.43.

Q.48. Write and explain various types of channels and channel modes used in GSM.

(R.G.P.V., Dec. 2009)

Ans. GSM specifies two basic groups of logical channels, i.e., traffic channels and control channels.

(i) **Traffic Channels (TCH)** – GSM uses traffic channels to transmit user data (e.g. voice, fax). Two basic categories of TCHs have been defined, i.e., full-rate TCH (TCH/F) and half-rate TCH (TCH/H). A TCH/F has a data rate of 22.8 kbit/s, whereas TCH/H only has 11.4 kbit/s. With the voice codecs available at the beginning of the GSM standardization, 13 kbit/s were required, whereas the remaining capacity of the TCH/F (22.8 kbit/s) was used for error correction (TCH/FS). Improved codes allow for better voice coding and can use a TCH/H. Using these TCH/HSs doubles the capacity of the GSM system for voice transmission. However, speech quality decrease with the use of TCH/HS and providers try to avoid using them. The standard codecs for voice are called full-rate (FR, 13 kbit/s) and half-rate (HR, 5.6 kbit/s). A newer codec; enhanced full-rate (EFR), provides better voice quality than FR as long as the transmission error rate is low. The generated data rate is only 12.2 kbit/s. New codecs, which automatically choose the best mode of operation depending on the error rate, i.e. AMR (adaptive multi-rate), will be used together with 3G systems. An additional increase in voice quality is provided by the so-called tandem free operation (TFO). This mode can be used if two MSs exchange voice data. In this case, coding to and from PCM encoded voice (standard in ISDN) can be skipped and the GSM encoded voice data is directly exchanged. Data transmission in GSM is possible at many different data rates, e.g., TCH/F4.8 for 4.8 kbit/s, TCH/F9.6 for 9.6 kbit/s, and as a newer specification, TCH/F14.4 for 14.4 kbit/s. These logical channels differ in term of their coding schemes and error correction capabilities.

(ii) **Control Channels (CCH)** – Many different CCHs are used in a GSM system to control medium access, allocation of traffic channels or mobility management. Three groups of control channels have been defined, each again with subchannels –

(a) **Broadcast Control Channel (BCCH)** – A BTS uses this channel to signal information to all MSs within a cell. Information transmitted in this channel is e.g. the cell identifier, options available within this cell (frequency hopping), and frequencies available inside the cell and in neighbouring cells. The BTS sends information for frequency correction via the frequency correction channel (FCCH) and information about time synchronization via the synchronization channel (SCH), where both channels are subchannels of the BCCH.

(b) **Common Control Channel (CCCH)** – All information regarding connection setup between MS and BS is exchanged via the CCCH. For calls toward an MS, the BTS uses the paging channel (PCH) for paging the appropriate MS. If an MS wants to setup a call, it uses the random access channel (RACH) to send data to the BTS. The RACH implements multiple

access (all MSs within a cell may access this channel) using slotted Aloha. This is where a collision may occur with other MSs in a GSM system. The BTS uses the access grant channel (AGCH) to signal an MS that it can use a TCH or SDCCH for further connection setup.

(c) **Dedicated Control Channel (DCCH)** – While the previous channels have all been uni-directional, the following channels are bi-directional. As long as an MS has not established a TCH with the BTS, it uses the stand-alone dedicated control channel (SDCCH) with a low data rate (782-bit/s) for signalling. This can comprise authentication, registration or other data needed for setting up a TCH. Each TCH and SDCCH has a slow associated dedicated control channel (SACCH) associated with it, which is used to exchange system information, such as the channel quality and signal power level. Finally, if more signalling information needs to be transmitted and a TCH already exists, GSM uses a fast associated dedicated control channel (FACCH). The FACCH uses the time slots which are otherwise used by the TCH. This is necessary in the case of handovers where BTS and MS have to exchange larger amounts of data in less time.

Channel Modes – The individual users can not have their own TCH at all times due to precious value of the radio spectrum.

(i) **Dedicated Mode** – Dedicated mode uses TCH during cell establishment and uses SACCH to perform location updating in the dedicated mode. SACCH and TCH are dedicated channels for both uplink and downlink channels.

(ii) **Idle Mode** – In the duration of non-call activities, the five downlink channels are in the idle mode – FCCH, SCH, BCCH, that is broadcasting regularly; PAGCH and CBCH, that transmits one message every 2 sec. In the duration of idle mode, the mobile station listens to the common downlink channels and also uses SDCCH to register a mobile location related to a particular base station to the network.

Q.49. Explain the protocol architecture of GSM.

Ans. Fig. 4.19 shows the protocol architecture of GSM with signalling protocols, interfaces. The main term to know about is the U_m interface, as the other interfaces occur between entities in a fixed network. **Layer 1**, the physical layer, handles all radio-specific functions. This includes the creation of bursts according to the five different formats, multiplexing of bursts into a TDMA frame, **synchronization** with the BTS, detection of idle channels, and measurement of the channel quality on the downlink. The physical layer at U_m uses GMSK for digital modulation and performs encryption/decryption of data, i.e., encryption is not performed end-to-end, but only between MS and BSS over the air interface.

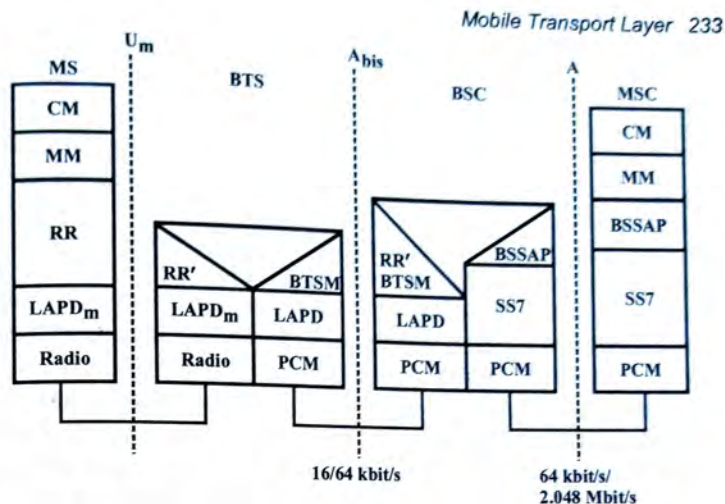


Fig. 4.19 Protocol Architecture for Signaling

Synchronization includes the correction of the individual path delay between an MS and the BTS. All MSs within a cell use the same BTS and hence must be synchronized to this BTS. The BTS produces the time structure of frames, slots etc. In this context, a problematic aspect are the different round trip times (RTT). An MS close to the BTS has a very short RTT, whereas an MS 35 km away already exhibits an RTT of around 0.23 ms. If the MS far away used the slot structure without correction, large guard space would be needed, as 0.23 ms are already 40 percent of the 0.577 ms available for each slot. So, the BTS transmits the current RTT to the MS, which then adjusts its access time so that all bursts reach the BTS within their limits. This mechanism decreases the guard space to only 30.5 μ sec or five percent. Adjusting the access is controlled via the variable timing advance, where a burst can be shifted to 63-bit times earlier, with each bit having a duration of 3.69 μ sec. If the available timing advance cannot be extended, a burst cannot be shifted earlier than 63-bit times. This results in the 35 km maximum distance between an MS and a BTS. It might be possible to receive the signals over long distances. Access cannot be permitted to avoid collisions at the BTS.

Signalling between entities in a GSM network requires higher layers. For this purpose, the LAPD_m protocol has been defined at the U_m interface for **layer 2**. LAPD_m, name derived from link access procedure for the D-channel (LAPD) in the ISDN systems, which is a version of HDLC. LAPD_m is a light weight LAPD because it does not need synchronization flags or checksumming for error detection. LAPD_m offers reliable data transfer over connections, re-sequencing of data frames, and flow control. As there is no buffering between

layer one and two, LAPD_m has to obey the frame structures, recurrence patterns etc. defined for the U_m interface. Further services provided by LAPD_m include segmentation and reassembly of data and acknowledged/unacknowledged data transfer.

Layer 3, network layer in GSM, comprises several sublayers as shown in fig. 4.19. The lowest sublayer is the radio resource management (RR). Only a part of this layer, RR', is implemented in the BTS, the remainder is situated in the BSC. The functions of RR' are supported by the BSC via the BTS management (BTSM). The main tasks of RR are setup, maintenance and release of radio channels. RR also directly accesses the physical layer for radio information and offers a reliable connection to the next higher layer.

Mobility management contains functions for registration, authentication, identification, location updating and the provision of a temporary mobile subscriber identity (TMSI) that replaces the international mobile subscriber identity (IMSI) and which hides the real identity of an MS user over the air interface. While the IMSI identifies a user, the TMSI is valid only in the current location area of a VLR. MM offers a reliable connection to the next higher layer.

Finally, the call management (CM) layer contains three entities – call control (CC), short message service (SMS) and supplementary service (SS). SMS allows for message transfer using the control channels SDCCH and SACCH. CC provides a point-to-point connection between two terminals and is used by higher layers for call establishment, call clearing and change of call parameters. This layer also provides functions to send-in-band tones, called dual tone multiple frequency (DTMF), over the GSM network.

Additional protocols are used at the A_{bis} and A interfaces. Data transmission at the physical layer typically uses pulse code modulation (PCM) systems. While PCM systems offer transparent 64 kbit/s channels, GSM also allows for the submultiplexing of four 16 kbit/s channels into a single 64 kbit/s channel. The physical layer at the A interface typically includes leased lines with 2.048 Mbit/s capacity. LAPD is used for layer two at A_{bis}, BTSM for BTS management.

Signalling system number 7 (SS7) is used for signalling between an MSC and a BSC. This protocol also transfers all management information between MSCs, HLR, VLRs, AuC, EIR and OMC. An MSC can also control a BSS via a BSS application part (BSSAP).

Q.50. Explain localization and calling in a GSM system.

Ans. Localization – The GSM system has one fundamental feature of automatic, worldwide localization of users. The system always knows where a

user currently is and the same phone number is valid worldwide. GSM performs periodic location updates to provide this service even if a user does not use the mobile station. The HLR always keeps information about the current location, and the VLR currently responsible for the MS informs the HLR about location changes. As soon as an MS moves into the range of a new VLR, the HLR transmits all user data required to the new VLR. Changing VLRs with uninterrupted availability of all services is also known as *roaming*. Roaming can occur within the network of one provider, between two providers in one country, but also between different providers in different countries. Several numbers are required to locate an MS and to address the MS.

(i) Mobile Station International ISDN Number (MS ISDN) – For a user of GSM, the only important number is the phone number. It should be recalled that the phone number is not associated with a certain device but with the SIM, that is personalized for a user. The MS ISDN follows the ITU-T standard E.164 for addresses as it is also used in fixed ISDN networks. This number consists of the country code (CC), the national destination code (NDC), and the subscriber number (SN).

(ii) International Mobile Subscriber Identity (IMSI) – For internal unique identification of a subscriber, GSM uses the IMSI, which consists of a mobile country code (MCC), the mobile network code (MNC), and finally the mobile subscriber identification number (MSIN).

(iii) Temporary Mobile Subscriber Identity (TMSI) – GSM uses the 4 byte TMSI for local subscriber identification. TMSI is chosen by the current VLR and is only valid temporarily and within the location area of the VLR. In addition, a VLR may change the TMSI periodically.

(iv) Mobile Station Roaming Number (MSRN) – Another temporary address that hides the identity and location of a subscriber is MSRN. The VLR produces this address on request from the MSC, and the address is also kept in the HLR. MSRN keeps the current visitor country code (VCC), the visitor national destination code (VNDC), the identification of the current MSC together with the subscriber number.

All these numbers described above are needed to find a subscriber and to maintain the connection with a mobile station.

Calling – Calling can be done by two types of call procedure i.e., mobile terminated call (MTC) and mobile originated call (MOC).

(i) Mobile Terminated Call (MTC) – It is a situation in which a station calls a mobile station. Fig. 4.20 shows the basic steps needed to connect the calling station with the mobile user.

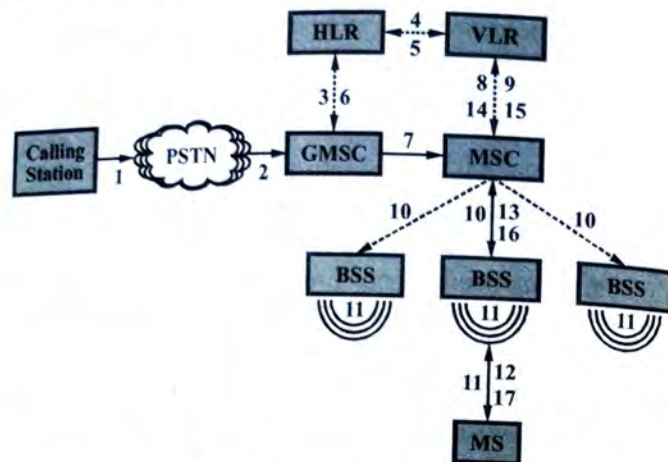


Fig. 4.20 Mobile Terminated Call (MTC)

- (1) A user dials the phone number of a GSM subscriber.
- (2) The fixed network PSTN (public switched telephone network) notices that the number belongs to a user in the GSM network and forwards the call setup to the Gateway MSC.
- (3) The GMSC identifies the HLR for the subscriber and signals the call setup to the HLR.
- (4) The HLR now checks whether the number exists and whether the user has subscribed the requested services, and requests an MSRN from the current VLR.
- (5) After receiving MSRN, VLR signals back to HLR.
- (6) Now, the HLR can determine the MSC responsible for the MS and forwards this information to the GMSC.
- (7) The GMSC can now forward the call setup request to the MSC.
- (8) Now, MSC is responsible for all further steps. First, it requests the current status of the MS from the VLR.
- (9) After checking it signals back to MSC.
- (10) If the MS is available, the MSC initiates paging in all cells it is responsible for.
- (11) The BTSs of all BSSs transmit this paging signal to the MS.
- (12 to 14) The MS answers, the VLR has to perform security checks.
- (15 to 17) The VLR then signals to the MSC to set-up a connection to MS.

(ii) **Mobile Originated Call (MOC)** – It is much simpler compared to MTC (see fig. 4.21). Following are the steps to made a call –

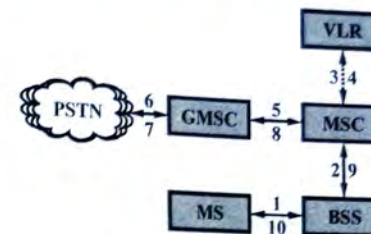


Fig. 4.21 Mobile Originated Call (MOC)

- (1) The MS transmits a request for a new connection.
- (2) The BSS forwards this request to the MSC.
- (3 to 6) The MSC checks if this user is allowed to setup a call with the requested service and checks the availability of resource through the GSM network and into the PSTN.
- (7 to 10) If all resources are available, the MSC set-up a connection between the MS and the fixed network.

Q.51. Differentiate HLR and VLR.

(R.G.P.V., Dec. 2015)

Ans. The HLR is the most important database software in a GSM system that controls the management of the mobile subscriber account. It contains subscriber's current location, service type, address and billing information. As soon as an mobile station leaves its current location area, the information in the HLR is updated. This information is necessary to localize a user in the worldwide GSM network.

The VLR associated to each MSC is a temporary database software. The VLR stores all important information about subscribers that are currently physically in the region covered by the switching center. If a new MS comes into an LA the VLR is responsible for, it copies all relevant information for this user from the HLR.

Both the HLR and the VLR are located in the network and switching subsystem (NSS).

Both the HLR and the VLR are needed in localization and calling. The HLR always contains information about the current location, and the VLR currently responsible for the MS informs the HLR about location changes. As soon as an MS moves into the range of a new VLR, the HLR sends all user data required to the new VLR. Changing VLRs with uninterrupted availability of all services is also called roaming.

Q.52. Write short note on handover.

Or

Define handover in GSM system.

Ans. Single cell do not cover the whole service area, but, e.g., only upto 35 km around each antenna on the country side and some hundred meters in cities. To cover whole area a cellular systems require a procedure called **handover**. The smaller the cell size and the faster the movement of a mobile station through the cells (upto 250 km/h for GSM), the more handovers of ongoing calls are required. However, a handover should not cause a cut-off, also called call drop. GSM aims at maximum handover duration of 60 ms.

Q.53. Explain the term handover. What are basic reasons for a handover?

Ans. Handover – Refer the ans. of Q.52.

Basic Reasons – There are two basic reasons for a handover –

(i) The mobile station moves out of the range of a BTS or a certain antenna of a BTS respectively. The received signal level decreases continuously until it falls below the minimal requirements for communication. The error rate may grow due to interference, the distance to the BTS may be too high (max. 35 km) etc. – all these effects may diminish the quality of the radio link and make radio transmission impossible in the near future.

(ii) The wired infrastructure (MSC, BSC) may decide that the traffic in one cell is too high and shift some MS to other cells with a lower load (if possible). Handover may be due to load balancing.

Q.54. Explain handover scenarios in GSM.

(R.G.P.V., Dec. 2012)

Ans. Fig. 4.22 shows four types of handover scenarios in GSM –

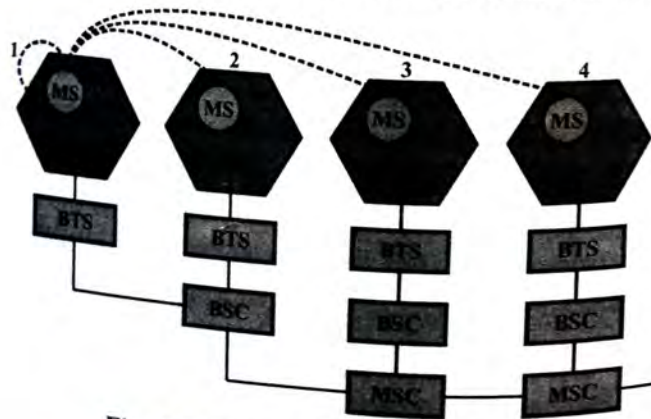


Fig. 4.22 Types of Handover in GSM

(i) **Intra-cell Handover** – Narrow-band interference could make transmission at a certain frequency impossible within a cell. The BSC could then decide to change the carrier frequency (scenario 1).

(ii) **Inter-cell, Intra-BSC Handover** – This is a typical handover scenario. The mobile station moves from one cell to another, but stays within the control of the same BSC. The BSC then performs a handover, assigns a new radio channel in the new cell and releases the old one (scenario 2).

(iii) **Inter-BSC, Intra-MSC Handover** – As a BSC only controls a limited number of cells, GSM also has to perform handovers between cells controlled by different BSCs. This handover then has to be controlled by the MSC (scenario 3).

(iv) **Inter-MSC Handover** – A handover could be required between two cells belonging to different MSCs. Now both MSCs perform the handover together (scenario 4).

Q.55. Explain handover techniques in cellular networks. What are the typical steps for handover, what types of handover can occur?

(R.G.P.V., June 2011)

Ans. Handover – Refer the ans. of Q.52.

The handover process can be divided into several subprocesses –

S.No.	Subprocess	Involved Network Element
(i)	Measurements – • DL level serving cell • DL quality serving cell • DL level neighbour cells • UL level • UL level quality • Timing advance	MS BTSE
(ii)	Measurement reporting and neighbour cell book keeping	MS BTSE
(iii)	Measurement preprocessing	BTSE
(iv)	Threshold comparison and handover detection	BTSE

(v)	Target cell list generation	BTSE
(vi)	Target cell list evaluation • intra BSS handover • inter BSS handover	BSC MSC
(vii)	Selection of new channel	BSC
(viii)	Handover execution	MS, STSE, BSC, MSC

Types of Handover – Refer the ans. of Q.54.

Q.56. Give reasons for a handoff in GSM and the problems associated with it. What types of handover can occur? (R.G.P.V., Dec. 2017)

Or

Give reasons for a handover in GSM and the problems associated with it. What types of handover can occur? (R.G.P.V., Nov. 2018)

Ans. Refer the ans. of Q.53 and Q.54.

Q.57. Explain the behaviour of the received signal level during handover.

Ans. Fig. 4.23 shows the typical behaviour of the received signal level while an MS moves away from one BTS (BTS_{old}) closer to another one (BTS_{new}). In this case, handover decision does not depend on the actual value of the received signal level, but on the average value. Therefore, the BSC collects all values from BTS and MS and calculates average values. These values are then compared to thresholds, i.e., the handover margin (HO_MARGIN), which includes some hysteresis to avoid a ping-pong effect

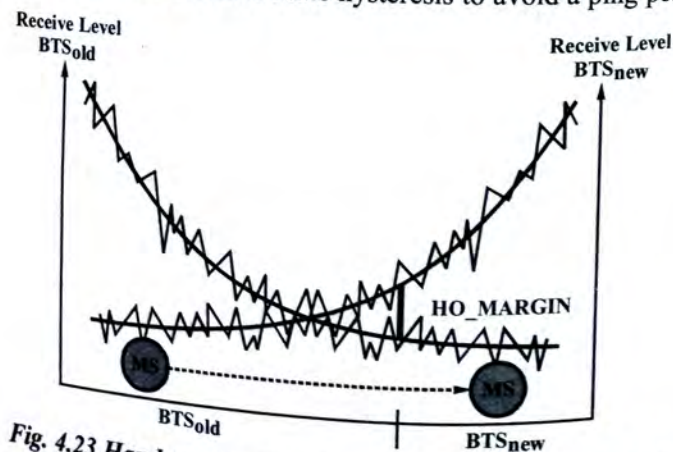


Fig. 4.23 Handover Decision Depending on Receive Level

(without hysteresis, even short term interference, e.g., shadowing due to a building, could cause a handover). Still, even with the HO_MARGIN, the ping-pong effect may occur in GSM – a value which is too high could cause a cutoff, and a value which is too low could cause too many handovers.

Q.58. Describe the process of call handover when a mobile station move. (R.G.P.V., Nov. 2019)

Ans. Handover is the procedure that transfers an ongoing call from one cell to another as the user's moves through the coverage area of cellular system. The purpose of the handover procedure is to preserve ongoing calls when the mobile station moving from one cell to another. In GSM measurements reports to perform handover, which is made by serving BSC which has no direct knowledge of the radio quality. These measurements reports contain the radio signal quality of the downlink from the BTS to MSC of the call and up to five neighbouring cells. The serving BTS measures the uplink from the MSC to BTS radio signal quality of the call and forward in the measurements reports. The information in the measurements reports the BSC is able to decide whether a handover to another cell is needed. These measurements reports are periodically transmitted from the MSC to BSC on the SACCH channel assigned to each communication for every connection. The repetition duration of the SACCH produces a fixed time grid of 480 ms in the measurements reports. The measured RXLEV_s from the BTS and from a neighbour to the measurements reports submitted during a call in many cases these measurements reports takes place to take handover procedure in a way to avoid shortcomings. A connection is continuously measured and evaluated by respective base station and MSC. Handover is based upon that evaluation. As the mobile users leaves the coverage area of the one base station must obtain coverage from the neighbouring station in order to keep the connection keep on. Cut off connection or drop call are not acceptable at any level during the call. Handover take places when the traffic level of cell reaches to high level or when neighbouring are being underutilized.

Q.59. Explain the security services included in GSM.

Ans. GSM offers several security services using confidential information stored in the AuC and in the individual SIM. The SIM stores personal, secret data and is protected with a PIN against unauthorized use. The security services offered by GSM are explained below –

(i) **Access Control and Authentication** – The first step includes the authentication of a valid user for the SIM. The user needs a secret PIN to access the SIM, the next step is the subscriber authentication.

(ii) **Confidentiality** – All user-related data is encrypted. After authentication, BTS and MS apply encryption to voice, data and signalling. This confidentiality exists only between MS and BTS, but it does not exist end-to-end or within the whole fixed GSM/telephone network.

(iii) **Anonymity** – To provide user anonymity, all data is encrypted before transmission and user identifiers (which would reveal an identity) are not used over the air. Instead, GSM transmits a temporary identifier (TMSI), which is newly assigned by the VLR after each location update. Additionally, the VLR can change the TMSI at any time.

Q.60. What are the functions of authentication and encryption in GSM? How is system security maintained in GSM?

(R.G.P.V., Dec. 2008, 2010, 2011)

Ans. Authentication – Authentication prevents unauthorized access in the network.

Phase I – The SIM is protected by a PIN (personal identification number) code. The SIM locally checks the PIN, hence the SIM is not sent out over the radio link.

Phase II – An inquiry is made by the GSM network using a random number (RAND). The 128-bit RAND and MS's secret parameter K_i is mixed in an A_3 processing algorithm, resulting in a 32-bit long SRES (signed result) number. This SRES is sent by MS to network for verification, as shown in fig. 4.24.

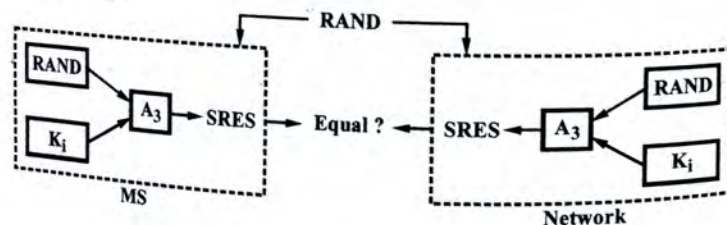


Fig. 4.24 The Computation of Authentication

Encryption – Encryption is used to prevent unauthorized listening. The RAND is sent from the network to the MS, and mixes K_i , is an A_8 processing algorithm, and results in a K_c (64-bits). The ciphering sequences are produced from the K_c , as shown in fig. 4.25. The K_c and frame number are input to a A_5 ciphering algorithm, and produces 114-bits, which uses an XOR operation between the 114-bits of plain text and ciphering sequence S_2 .

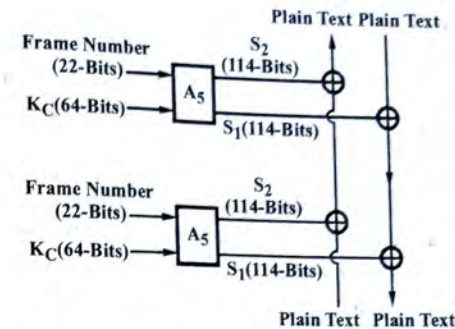


Fig. 4.25 Ciphering and Deciphering

Security Management – The repositories of the subscriber's key K_i are SIM (MS side) and AuC (network side). The K_i does not transmit over the air. Both sides use A_3 and A_8 algorithm.

Q.61. Write short note on roaming.

Ans. Roaming refers to a wireless network service extension in an area that differs from the registered home network location. Roaming enables a mobile device to access the Internet and other mobile services when out of its normal coverage area. It also gives a mobile device the ability to move from one access point to another.

Roaming is derived from real-time optimally adapting mesh (ROAM).

Roaming services are usually provided by cellular service providers as well as Internet service providers (ISP) via a cooperative agreement. Traditional cellular roaming services are provided by both the Global System for Mobile Communications (GSM) and code division multiple access (CDMA) operators. Services are either free or billed according to local area rates. Wireless telecommunication roaming services are usually included in mobile/cellphone subscriber service packages for use outside local networking zones.

GSM/WLAN roaming services can be supplied in two different scenarios. One is SIM-based roaming and the second is username/password base roaming.

Wireless local area network (WLAN) roaming services are segmented as follows –

(i) **Internal Roaming** – Implemented when a mobile station is transferred with a strong signal between access points, preventing network blockage or interruption from weak signals.

(ii) **External Roaming** – Implemented when a mobile station shifts to a wireless LAN or other foreign Wireless Internet Service Provider (WISP) to access service. WISP allows users to maintain an Internet connection while moving within a local coverage area.

An ISP uses special software to automatically track roaming usage and corresponding billing. To benefit from roaming, subscribers should have an ISP connection that supports roaming. A traveling user may route calls to the ISP's locally assigned number after logging into a foreign ISP through a computer modem. The foreign ISP provides Internet access after validating the user's home mail server.



UNIT

5

MOBILE DEVICE OPERATING SYSTEMS

SPECIAL CONSTRAINTS & REQUIREMENTS, COMMERCIAL MOBILE OPERATING SYSTEMS, SOFTWARE DEVELOPMENT KIT – IOS, ANDROID, ETC.

Q.1. Discuss the special constraints and requirements of mobile operating system (O/S).

Ans. The special constraints and requirements of mobile operating system are as follows –

(i) Limited Screen Size –

(a) Size of mobile must be small and portable screen size is important.

(b) New innovative user interface is needed.

(c) Switching between menu and iconic display.

(ii) Limited Processing Power –

(a) ARM – Based processor.

(b) It is energy efficient, cheaper, powerful.

(c) It is slower processor.

(d) Size of on – chip memory is restricted.

(e) So the development is carried out outside.

(iii) Limited and Fluctuating Bandwidth of the Wireless Medium –

(a) Mobile need to run complex protocols caused by mobility and wireless medium.

(b) Medium is directly prone to noise leads to bit-errors rates.

(c) Bandwidth may fluctuate due to noise, mobility of the node and obstacles.

(d) This show up with short-term fades.

(e) Longer disconnections due to handoffs.

(f) Uninterrupted communication requires special support like data caching, pre-fetching and integration.

(iv) Limited Memory –

- (a) Less amount and both storage and volatile memory.
- (b) OS must small as much as possible.
- (c) Provide rich user functionalities when user demands.
- (d) Size of the kernel plays very vital role in this.

(v) Miniature KeyBoard –

- (a) Use of small size keyboard and stylus.
- (b) Typing is too difficult for large document creation.
 - (1) We need auto completion option.
 - (2) Free from writing and handwriting recognition.

(vi) Limited Battery Power –

- (a) Mobile need to lightweight and portable so we need to use slim batteries.
- (b) No frequent recharges.
- (c) OS needs to be computationally efficient and minimum power consumption.
- (d) Putting processor and display into sleep mode when it is not needed.

(vii) Special Requirements –

- (a) Support for specific communication protocols.
- (b) Support for variety of input mechanisms.
- (c) Compliance with open standards.
- (d) Extensive library support.

Q.2. Discuss various types of commercial mobile operating system.

Ans. Various types of commercial mobile operating system are as follows-

IBM Simon –

- (i) Multifunctional phone
- (ii) OS is ROM-DOS with file staker
- (iii) Features –
 - (a) E-mails
 - (b) Fax
 - (c) Cellular pages

(iv) Applications –

- (a) Addressbook
- (b) Calendar
- (c) Appointment scheduler
- (d) World time clock
- (e) Electronic note pad
- (f) Stylus input for keyboard.

Palm OS (Palm Pilot 1000)

(i) It is PDA and Phone that uses the motorola processor of 16 Mhz with 256 kb or 512 kb of built in RAM.

(ii) It uses the palm OS.

(iii) Features –

- (a) Monochrome LCD panels with handwriting recognition.

(iv) Applications –

- (a) Date book
- (b) Address book
- (c) To do list
- (d) Memo pad.

Palm OS

(i) Simple, single-tasking environment to allow launching of full screen applications with a basic, common GUI set.

(ii) Handwriting recognition input system called Graffiti 2.

(iii) Sound playback and record capabilities.

(iv) TCP/IP network access

(v) Expansion memory card support.

(vi) Monochrome or colour screens with resolutions up to 480×320 pixel.

(vii) HotSync technology for data synchronization with desktop computers.

(viii) Simple security model – Device can be locked by password, arbitrary application records can be made private.

(ix) Serial port/USB, infrared, bluetooth and Wi-Fi connections.

(x) Defined standard data format for personal information management applications to store calendar, address, task and note entries, accessible by third-party applications.

Windows Embedded Compact (Windows CE)

- (i) Operating system developed by Microsoft.
- (ii) Optimized kernel runs on 1 MB of memory.
- (iii) It uses the hybrid kernel.
- (iv) Platforms –
 - (a) AutoPC
 - (b) Pocket PC 2000
 - (c) Pocket PC 2002
 - (d) Windows Mobile 2003
 - (e) Windows Mobile 2003 SE
 - (f) Windows Mobile 5
 - (g) Windows Mobile 6
 - (h) Smartphone 2002
 - (i) Smartphone 2003
 - (j) Portable media center
 - (k) Zune and now Windows phone.

Symbian OS

- (i) It supports communication protocols like TCP, UDP, PPP, DNS, FTP, WAP.
- (ii) Low power mode of CPU switch available.
- (iii) Event based applications run by active objects.
- (iv) Eclipse plug-in is available for development.
- (v) It supports bluetooth, infraRed, USB connectivity.
- (vi) Low on power and memory requirement applications (OOPS) based.
- (vii) Carbide is an IDE supports for C++ development.

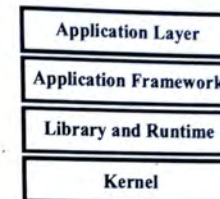
Android OS

- (i) Google's operating system.
- (ii) Open source
- (iii) Features –
 - (a) Phone based keyboard or touchscreen for input.
 - (b) Has built-in web browser that renders the web page fully.
 - (c) Easy way to develop third party applications.
 - (d) Android SDK and Eclipse.

- (e) Provides RDBMS SQLite for data storage.
- (f) Pre installed applications such as Gmail, maps, voice-search, etc.

(iv) Application layer –

- (a) Web browser, e-mail client, SMS manager, maps.
- (b) Apps written in J2ME.
- (c) No priority control to manage resources aggressively.



(v) Application framework –

- (a) Used to implement standard structure for apps.
- (b) It provides the set of services that can used by developers.
- (c) Services –
 - (1) Managers (for events)
 - (2) Content providers (for accessing data)

(vi) Libraries and Runtime –

- (a) It is written in multiple languages including C and C++ and called through JAVA interfaces.
- (b) It includes surface managers (for compositing windows) 2D and 3D graphics, media codecs (MP4, MP3) and SQL database SQLite.
- (c) It includes web browser engine called webkit.
- (d) Runtime –
 - (1) Core libraries
 - (2) Dalvik virtual machine
- (e) DVM converts the java code into native ARM code.
- (f) Each application runs as own process with own instance of Dalvik virtual machine (DVM).

(vii) Kernel

- (a) It is based on Linux kernel.
- (b) Excluded native X windows of GNU.

- (c) Android implements it's own –
 - (1) Device drivers
 - (2) Memory management
 - (3) Process management
 - (4) Networking functionalities.
- (d) Runs the application concurrently
- (e) Google initially maintains the kernel code and contributed to linux public distribution.
- (f) Now google no longer maintains the kernel extensions.

iOS (Apple)

- (i) Apple developed iOS iPhone the direct replacement for iPod.
- (ii) It is property and maintained apple.
- (iii) It is derivative of Mac OS.
- (iv) Market shots
 - (a) Geatures
 - (b) Swipe
 - (c) Tap
 - (d) Pinch
 - (e) Reverse pinch
 - (f) It includes internal accelerometer for recognize the shake and change the music.
 - (g) Display mode switch portrait to landscape and vice versa.

Blackberry OS

- (i) It is a proprietary operating system designed for blackberry smartphones produced by research in motion limited (RIM).
- (ii) It has a very good e-mail system at user level.
- (iii) Architecture of blackberry OS is not revealed.
- (iv) It supports instant mailing while maintaining high level of security through on-device hardware based message encryption.

Q.3. Discuss the software development kit in brief.

Ans. A software development kit (SDK or devkit) is typically a set of software development tools that allows the creation of applications for a certain software package, software framework, hardware platform, computer system, video game console, operating-system, or similar development platform.

iOS Software Development Kit –

- (i) Developed by apple Inc.
- (ii) Package contains
 - (a) Development IDE
 - (b) Iphone Simulator
 - (c) Applications written in swift and objective – C.
 - (d) Some elements of application developed using C and C++.
- (iii) Media
 - (a) OpenAL
 - (b) Audio mixing and recording
 - (c) Video playback
 - (d) Image file formats
 - (e) Quartz
 - (f) Core animation
 - (g) OpenGL ES
- (iv) Mac OS X kernel
 - (a) TCP/IP
 - (b) Sockets
 - (c) Power management
 - (d) File system
 - (e) Security
- (v) Cocoa touch
 - (a) Multi-touch events and controls
 - (b) Accelerometer support
 - (c) View hierarchy
 - (d) Localization (i18n)
 - (e) Camera support
- (vi) Core services
 - (a) Networking
 - (b) Embedded SQLite database
 - (c) Core location
 - (d) Threads
 - (e) CoreMotion

Android Software Development Kit

- (i) Package contains
 - (a) Android SDK, IDE for development (eclipse or android studio).
 - (b) Virtual machine manager and system images to boot the emulators.
 - (c) Tools to debug the apps during developments.
 - (d) Builders for development (ant for eclipse and gradle for android studio).
- (ii) Android application components
 - (a) Activity
 - (b) Services
 - (c) Broadcast receivers.

M-COMMERCE – STRUCTURE, PROS & CONS, MOBILE PAYMENT SYSTEM, SECURITY ISSUES

Q.4. What is m-commerce ? Give its applications.

Ans. M-commerce (mobile commerce) is the buying and selling of goods and services through wireless handheld devices such as cellular telephone and personal digital assistants (PDAs). Known as next-generation e-commerce, m-commerce enables users to access the Internet without needing to find a place to plug in.

M-commerce applications can be broadly categorized into two –

(i) **Business-to-Consumer Applications (B2C)** – Business-to-consumer is a form of commerce in which products or services are sold by a business firm to a consumer.

- (a) Advertising
- (b) Comparison shopping
- (c) Information about a product
- (d) Mobile ticketing
 - (1) We can buy M-Tickets using credit cards.
 - (2) Purchase confirmation can be sent through SMS or e-mail.
 - (3) Train, movie, bus, etc.

- (e) Loyalty and payment services
 - (1) Payback cards.
 - (2) Points will be generated according to the user buy's.
 - (3) According to points user earns they may avail special gifts.
 - (f) Interactive advertisements
 - (1) Offers through the TV and Teleshopping
 - (g) Catalogue shopping
 - (1) Direct buy using direct link sent by company.
- (ii) **Business-to-business Applications (B2B)** –
- (a) Ordering and delivery confirmation
 - (1) Mobile phones can be used by the dealers to order products.
 - (2) Mobile phones can be used to gather information about the status of consignments during the transport and delivery process.
 - (3) Realtime consignment tracking.
 - (b) Stock tracking and control
 - (1) Mobile tracking can be connected with inventory of store
 - (2) It can track warehouse status
 - (c) Mobile inventory management
 - (d) Supply chain management (SCM).

Q.5. Draw and explain the structure of m-commerce.

Ans. The architecture of mobile commerce application is shown in fig. 5.1.

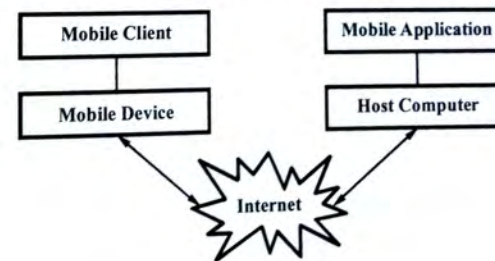


Fig. 5.1

- (i) Normally mobile applications consists of
 - (a) Client-side programs gives the nice user interface to the user.
 - (b) Server-side programs gives
 - (1) Accessing the database
 - (2) Computations.
- Server-side programs resides on remote host computers.

(ii) Mobile devices

(a) It provides the user interface to user and moves the data from device to Internet and vice versa.

(b) Feature required.

- (1) Good Internet connectivity
- (2) Ability to display rich content like images
- (3) Equipped with good quality of camera with autofocus
- (4) Screen should display the bar codes properly
- (5) Ability to read RFID tags
- (6) SMS, MMS capability to send and receive
- (7) Ability to communicate between mobile devices and supporting networks
- (8) Ability to scan the bar codes
- (9) Ability to interact with the point-of-sale (PoS) terminals.

(iii) Mobile middleware

- (a) It is used to map the Internet content to mobile phones
- (b) It mostly handles the encryption and decryption in communication

(c) It provides the secure transactions.

(iv) Network

- (a) It gives the access to the mobile through the wireless medium
 - (b) Clients are connected through wireless networks
 - (c) Host is connected to wired networks
 - (d) Both are have the access to basic security functionalities.
- (v) Host computers
- (a) It is servers that executes the server side mobile applications.
 - (b) It consists of, web servers, database servers, applications programs and support softwares.

Q.6. Write down the advantages and disadvantages of m-commerce.

Ans. Advantages of M-commerce are given below –

- (i) For business organization, the benefits of using m-commerce include,
 - (a) Customer convenience
 - (b) Cost savings
 - (c) New business opportunities.

(ii) From the customer's perspective,

(a) M-commerce provides the flexibility of anytime, anywhere shopping using lightweight devices.

(b) The customer can save substantial time compared to visiting several stores for identifying the right product at the lowest price.

(iii) Mobile devices can be highly personalized, thereby providing an additional level of convenience to the customers.

(a) Repeated orders can be placed at the touch of a button.

Disadvantages of m-commerce are given below –

(i) Mobile devices do not generally offer graphics or processing power of a PC. The users therefore constrained to use small screen and keyboard and low resolution pictures and videos.

(ii) The small screens of mobile devices limit the complexity of applications. For example, the menu choice, and text typing capability are severely constrained.

(iii) The underlying network imposes several types of restrictions. For example, the available bandwidth is severely restricted, and international reach is prohibitively expensive. Therefore, ubiquity of m-commerce is hard to achieve in practice.

Q.7. What is mobile payment system ? Give its various schemes.

Ans. Mobile payment (M-payment) is defined as any payment instrument where a mobile device is used to initiate, authorize and confirm an exchange of financial value in return for goods and services.

Mobile Payment Schemes – Three popular types of M-payment schemes are currently being used –

(i) Bank account based M-payment

(a) The bank account number is linked to customer's mobile number.

(b) Customer may make a transactions with vendor, based on the bluetooth or wireless LAN connectivity with vendor.

(c) The bank account of the customer is debited and the value is credited to the vendor's account.

(d) M-check is linking credit or debit card with customer's mobile number.

(ii) Credit card based M-payment

(a) The credit card number is linked with mobile number.

(b) When a customer makes a transactions with merchant, credit card is charged and the value is added merchant's account.

(iii) Micropayment

(a) It is intended for payment for small purchases such as vending machines.

(b) Mobile can connect with bluetooth and wireless LAN connect to make payment.

Each payment scheme uses customer's banking information the service provider may charge small amount.

Q.8. Give the list of m-commerce security issues.

Ans. M-commerce security issues are given below –

- (i) Privacy risks.
- (ii) Mobile devices difficult to find on the move.
- (iii) Mobile devices go online and offline frequently.
- (iv) Attacks would be very difficult to trace.
- (v) Risk of mobile loss and theft.
- (vi) Fraud payment from stolen mobile is more difficult to track.
- (vii) Lack of any satisfactory mechanism to authenticate a particular user.



RGPV

CS/IT-8401
B.E. (Eighth Semester)
EXAMINATION, June 2010
(Common for CS & IT Engg. Branch)
WIRELESS NETWORK AND COMPUTING
(Elective-IV)

Note : (i) Attempt any five questions.

(ii) All questions carry equal marks.

1. (a) Write a detailed note on mobile Ad-hoc networks. (See Unit-III, Page 128, Q.13)
(b) Explain snooping TCP. What are its advantages and disadvantages ? (See Unit-IV, Page 190, Q.8)
2. (a) Explain in detail about context aware computing. **
(b) Explain Bluetooth protocol and protocol stack. **
3. With necessary diagrams explain the different types of Random Access methods of mobile data services. **
4. (a) Explain the overview of the IEEE 802.11. (See Unit-III, Page 134, Q.19)
(b) With necessary diagrams explain the HIPERLAN-1 requirements and architecture. **
5. (a) How handover takes place in indirect TCP ? (See Unit-IV, Page 188, Q.4)
(b) Explain query processing of data. **
6. What are the requirements associated with the mobile IP standard ? How they are met by mobile IP ? (See Unit-III, Page 166, Q.47)
7. (a) Explain the errors in wireless network, which degrade TCP performance ? **
(b) What are the steps involved in retrieving the indexed data frames ? **
8. Write short notes on the following : **
(a) DM **
(b) Linear predictive coding
(c) Media access control. (See Unit-III, Page 125, Q.8)

RGPV

IT-702(N)
B.E. (Seventh Semester)
EXAMINATION, Dec. 2010
(New Scheme)
(Information Technology Engg. Branch)
WIRELESS AND MOBILE COMPUTING

Note : All questions are compulsory.

**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus

1. (a) What is fading ? Write difference between flat and selective fading. 10
 (b) What are the advantages and disadvantages of CDMA for a cellular network ? ** 10

Or

2. (a) Draw and explain the model of wireless digital communication system. ** 10
 (b) What is the main physical reason for the failure of many MAC schemes known from wired networks ? What is done in wired network to avoid this effect ? (See Unit-III, Page 128, Q.12) 10
3. (a) Name the main elements of the GSM system architecture and describe their functions. (See Unit-IV, Page 227, Q.43) 10
 (b) What are the functions of authentication and encryption in GSM ? How is system security maintained ? (See Unit-IV, Page 242, Q.60) 10

Or

4. Write short notes on the following : 20
 (a) DECT **
 (b) UMTS **
 (c) TETRA **
 (d) Handover. (See Unit-IV, Page 238, Q.52) **
5. (a) Write advantages and disadvantages of WLANs. 10
 (b) Draw and explain protocol architecture and bridging for IEEE 802.11. 10
 (See Unit-III, Page 119, Q.3) (See Unit-III, Page 141, Q.30)

Or

6. (a) Compare wireless network standards IEEE 802.11 b, IEEE 802.11 a and HiperLAN2. ** 10
 (b) Write a short note on Bluetooth. ** 10
7. (a) Explain Mobile IP in terms of goals, assumptions and requirements. 10
 (See Unit-III, Page 167, Q.48)
 (b) How can DHCP be used for mobility and support of mobile IP ? 10
 **

Or

8. (a) Name the main differences between multi-hop ad-hoc networks and other networks. What advantages do these ad-hoc networks offer ? 10
 (See Unit-III, Page 133, Q.18)
 (b) What are the differences between AODV and the standard distance vector algorithm ? (See Unit-III, Page 178, Q.59) 10
 What are extensions needed ?

**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus

9. (a) List three design goals of a firewall. (See Unit-I, Page 47, Q.50) 10
 (b) Write a short note on Trojan horse defense. ** 10

Or

10. (a) What are three benefits that can be provided by an intrusion detection system ? ** 10
 (b) What is a honeypot ? ** 10

RGPV

IT-702(N)
 B.E. (Seventh Semester)
 EXAMINATION, June 2011
 (Information Technology Engg. Branch)
 WIRELESS AND MOBILE COMPUTING

Note : Attempt any five questions. All questions carry equal marks.

1. (a) Explain handover techniques in cellular networks. What are the typical steps for handover, what types of handover can occur ? 10
 (See Unit-IV, Page 239, Q.55)
 (b) Define Antenna. Explain various types of antennas. ** 10
2. (a) Explain different multiple access techniques. ** 10
 (b) Explain various propagation modes. ** 5
 (c) What are the differences between fast fading and slow fading ? 5
 **
3. (a) Explain the architecture of GSM. Which types of different services does GSM offer ? Give examples. (See Unit-IV, Page 230, Q.46) 10
 (b) Why and when are different signalling channels needed ? What are the differences ? ** 10
4. (a) With neat sketch, explain architecture of 802.11 LAN. Also explain its MAC logic. (See Unit-III, Page 134, Q.20) 10
 (b) Show the interaction of mobile-IP with standard TCP. Draw the packet flow from a fixed host to a mobile host via a foreign agent, then a handover takes place. What are following action of mobile IP and how does TCP react ? (See Unit-IV, Page 186, Q.3) 10
5. (a) What is Mobile-IP ? Explain the basic operation of Mobile-IP. 10
 (See Unit-III, Page 163, Q.45)
 (b) Explain the architecture of Bluetooth system. What are the Bluetooth technological characteristics ? ** 10
6. (a) What is DHCP ? How does DHCP support dynamic address allocation ? ** 10
 (b) What is Ad-hoc Network ? What are the characteristics of it ? 10
 (See Unit-III, Page 128, Q.13)

**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus

7. (a) Explain the threats and security issues to wireless networks. 10
 (b) What is Intruder ? Explain Intrusion detection models. **
 (c) What is firewall ? Explain design principles of firewall. ** 5

8. Write short notes on any four of the following – (See Unit-I, Page 247, Q.50) 5

- (i) DECT **
 (ii) GPRS 5 each
 (iii) WAP **
 (iv) HIPERLAN **
 (v) HSCSD **
 (vi) Snooping TCP. **

(See Unit-IV, Page 189, Q.6)

RGPV

IT-702
 B.E. (Seventh Semester)
 EXAMINATION, Dec. 2011
 (Information Technology Engg. Branch)
 WIRELESS AND MOBILE COMPUTING

Note : All questions are compulsory. All questions carry equal marks.

1. What are main benefits of a spread Spectrum system ? Explain Direct sequence spread spectrum and Frequency hopping spread spectrum. 20
 (See Unit-III, Page 146, Q.35)

Or

2. (a) Explain the following problems – 10
 (i) Hidden and Exposed terminals problem
 (ii) Near and Far terminals problem.

(See Unit-III, Page 126, Q.9)

- (b) Explain the term interference in the space, time, frequency and code domain. What are counter-measures in SDMA, TDMA, FDMA and CDMA systems. ** 10

3. (a) Draw and explain functional architecture of GSM system. 10
 (See Unit-IV, Page 227, Q.43)
 (b) Explain the term handover. What are basic reason for a handover ?
 (See Unit-IV, Page 238, Q.53) 10

Or

4. (a) What is GPRS ? Explain its architecture. ** 10
 (b) Write short notes on the following –

**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus

Wireless and Mobile Computing (VII-Sem.)

- (i) DECT **
 (ii) TETRA. **

5. (a) Explain various IEEE 802 standards and basic architecture of IEEE 802.11 standard. (See Unit-III, Page 141, Q.28) 10
 (b) Explain the functions of various layers of WLAN and the protocol architecture. (See Unit-III, Page 142, Q.32) 10

Or

6. (a) Compare IEEE 802.11, HIPERLAN-2 and Bluetooth with regard to their ad-hoc capabilities. ** 10
 (b) How do IEEE 802.11, HIPERLAN-2 and Bluetooth respectively, solve the hidden terminal problem. ** 10
 7. (a) Name the requirements for a mobile IP and justify them. Does mobile IP fulfill them all. (See Unit-III, Page 166, Q.47) 10
 (b) What is the basic purpose of DHCP ? Name the entities of DHCP. ** 10

Or

8. (a) Why is routing in multi-hop ad-hoc network complicated ? What are the special challenges ? (See Unit-III, Page 177, Q.58) 10
 (b) How does Dynamic source routing handle routing ? 10
 (See Unit-III, Page 179, Q.61)
 9. (a) List and briefly define three classes of intruders. ** 10
 (b) What are two common techniques used to protect a password file ? ** 10

Or

10. (a) Write a short note on intrusion detection system. ** 10
 (b) What are the functions of authentication and encryption in GSM ? How is system security maintained ? (See Unit-IV, Page 242, Q.60) 10

RGPV

IT-702
 B.E. (Seventh Semester)
 EXAMINATION, Dec. 2012
 WIRELESS AND MOBILE COMPUTING

- Note :** (i) Attempt any five questions (one question from each unit).
 (ii) All question carry equal marks.
 (iii) Assume suitable data if required.

Unit-I

1. (a) Explain following terms – 10
 (i) Handoff (See Unit-IV, Page 219, Q.35)

**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus

- (ii) GSM
(iii) OSI layer.

(b) What is mobile computing? Explain application of mobile computing. **
10

Or

2. (a) In time division multiple access, explain demand assignment multiple Access with explicit reservation and with implicit reservation. 10
**
(b) Compare infrared and radio transmission and explain the architecture of a cellular system. 10
(See Unit-IV, Page 203, Q.24)

Unit-II

3. (a) Explain GPRS architecture reference model? ** 10
(b) What is the function of following switches and data bases in network and switching sub-system of GSMs – 10
(i) Mobile services switching center
(ii) Home location register.

(See Unit-IV, Page 230, Q.45)

Or

4. (a) What is mobility management, DECT? Explain. ** 10
(b) Explain handover scenarios in GSM. 10
(See Unit-IV, Page 238, Q.54)

Unit-III

5. (a) Explain the three classes of transactional service provided by wireless transaction protocol. ** 10
(b) Why indirect TCP is developed? How packet loss is handled in case of fixed network to mobile host and mobile host to fixed networks. How does I-TCP transaction protocol work? 10

(See Unit-IV, Page 186, Q.3)

Or

6. (a) Describe 802.11 architecture. How can you achieve wireless LAN security? 10
(See Unit-III, Page 137, Q.22)
(b) Explain WAP push Architecture. How many operations are offered by push access protocol? ** 10

Unit-IV

7. (a) Explain the agent advertisement method of mobile-IP. What are the advantages and disadvantages of snooping TCP? 10
(See Unit-IV, Page 191, Q.9)
(b) Explain dynamic host configuration protocol. ** 10

Or

**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus

8. (a) Explain mobile IP registration process. How mobile TCP different from normal TCP? 10
(See Unit-III, Page 173, Q.50)
(b) If 20 MHz of total spectrum is allocated for a duplex wireless cellular system and each simplex channel has 25 kHz RF Bandwidth find : 10
(i) The no. of duplex channels
(ii) The total No. channel per cell site
If $N = 4$ cell reuse is used.

(See Unit-IV, Page 194, Prob.1)

Unit-V

9. (a) How does the biometrics authentication system work. Explain. 10
**
(b) Write a short note on – 10
(i) Intruders **
(ii) Virus **
(iii) Threats **
(iv) Worms **
(v) Trojan Horse. **

Or

10. (a) What is firewall design principle? How it work against attacks and threats? 10
(See Unit-I, Page 49, Q.53)
(b) What is password management? Explain the different kinds of password management algorithms. ** 10

IT-702
B.E. (Seventh Semester)
EXAMINATION, Dec. 2013
WIRELESS AND MOBILE COMPUTING

- Note :** (i) Attempt any *one* questions from each unit.
(ii) All question carry equal marks.

Unit-I

1. (a) Discuss the similarities and differences between a conventional cellular radio system and a satellite cellular radio system. ** 7
(b) Explain the cellular design techniques that are needed to improve coverage and capacity in cellular systems. 7
(See Unit-IV, Page 209, Q.27)

Or

2. (a) Discuss the various effects of mobile radio propagation. 7
(See Unit-IV, Page 215, Q.32)

**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus

- (b) Compare between SDMA, TDMA, FDMA and CDMA used in wireless networks. ** 7

Unit-II

3. (a) Discuss the different types of mobile services provided by GSM. Explain each in brief. (See Unit-IV, Page 225, Q.42) 7

- (b) Explain the architecture of GPRS. ** 7

Or

4. (a) What is mobility management? Discuss its various functions. 7

- (b) Explain briefly – **

(i) TETRA 7

(ii) UMTS. **

Unit-III

5. (a) Describe the physical layer architecture of 802.11 standard. 7

(See Unit-III, Page 143, Q.33)

- (b) What are the advantages and disadvantages of using Bluetooth based devices as a sensor network? Explain. ** 7

Or

6. (a) Draw and explain HIPERLAN-2 protocol stack. ** 7

- (b) What will be the impact on piconet in Bluetooth devices are connected to mobile units? Explain. ** 7

Unit-IV

7. (a) Explain how routing is performed on ad-hoc networks. 7

(See Unit-III, Page 174, Q.54)

- (b) What do you understand by mobile IP? Discuss its basic operation. (See Unit-III, Page 163, Q.45) 7

Or

8. (a) What is DHCP? How does DHCP support dynamic address allocation? ** 7

- (b) Write short notes –

(i) Snooping TCP

(ii) Mobile TCP. (See Unit-IV, Page 189, Q.6)

(See Unit-IV, Page 191, Q.10)

Unit-V

9. (a) Differentiate between viruses and related threats and worms. ** 7

- (b) What is firewall? Explain the design principle of firewall. 7

(See Unit-I, Page 247, Q.50)

Or

**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus

10. Write short notes –

(i) Intruders

(ii) Password management **

(iii) Trojan horse **

(iv) Biometrics. **

**

RGPV**IT-702****B.E. (Seventh Semester)****EXAMINATION, Dec. 2014****WIRELESS AND MOBILE COMPUTING**

Note : Attempt all questions. With internal choice.

Unit-I

1. (a) State some feature, application, of CDMA. ** 7

- (b) Define efficiency of TDMA. ** 7

Or

2. (a) Explain different type of fading model. ** 7

- (b) Compare SDMA, TDMA, FDMA, CDMA. ** 7

Unit-II

3. (a) What is meant by 'capacity on demand' principle in GPRS network? **

- (b) Distinguish between soft handoff and hard handoff. **

- (c) What are the types of GSM services? Explain. 14

(See Unit-IV, Page 225, Q.42)

Or

4. Draw the GSM architecture and explain the call establishment in GSM using the logical channel. (See Unit-IV, Page 229, Q.44) 14

Unit-III

5. (a) Draw and explain the frame format of MAC management. 7

(See Unit-III, Page 159, Q.42)

- (b) Explain HIPERLAN in detail. ** 7

Or

6. (a) Explain the architecture of IEEE 802.11. (See Unit-III, Page 137, Q.21) 7

- (b) Explain the MAC management sublayer in terms of handoff and power management. (See Unit-III, Page 160, Q.43) 7

Unit-IV

7. (a) What are the requirement of mobile IP? (See Unit-III, Page 166, Q.47) 7

- (b) Explain Ad-hoc network in terms of characteristic and performance (See Unit-III, Page 175, Q.55) 7

**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus

Or

8. Explain following term –

- (i) Snooping TCP
- (ii) Indirect TCP.

(See Unit-IV, Page 189, Q.6)
(See Unit-IV, Page 186, Q.3)

Unit-V

9. Explain in detail different biometric and authentication system.

Or

10. (a) Explain different types of firewall and discuss design principle of firewall.
(b) Define intrusion detection system.

(See Unit-I, Page 49, Q.54) 10
** 4

RGPV**IT-702**

B.E. (Seventh Semester)
EXAMINATION, Dec. 2015
WIRELESS AND MOBILE COMPUTING

- Note :** (i) Answer five questions. In each question part A, B, C is compulsory and D part has internal choice.
(ii) All parts of each questions are to be attempted at one place.
(iii) All questions carry equal marks, out of which part A and B (Max. 50 words) carry 2 marks, part C (Max. 100 words) carry 3 marks, part D (Max. 400 words) carry 7 marks.
(iv) Except numericals, Derivation, Design and Drawing etc.

Unit-I

1. (a) What is isotropic radiator ? Draw radiation pattern of isotropic radiator.
(b) What is fading ? Define its types.
(c) Define types of radio propagation mechanisms.
(d) What is the process of handoff ? Define types of handoff.

(See Unit-IV, Page 215, Q.33)

(See Unit-IV, Page 217, Q.34)

Or

Explain multiple access techniques such as SDMA, TDMA, FDMA and CDMA.

Unit-II

2. (a) Differentiate HLR and VLR.
(b) Define handover in GSM system.
(c) List the important GSM air-interface standard specifications.

(See Unit-IV, Page 237, Q.51)

(See Unit-IV, Page 238, Q.52)

(See Unit-IV, Page 224, Q.41)

**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus

Wireless and Mobile Computing (VII-Sem.)
Draw and explain the GSM network architecture.

(See Unit-IV, Page 227, Q.43)

Or

Explain the architecture of GPRS system. What are the enhancements over GSM ?

**

Unit-III

3. (a) Write WLAN advantages and disadvantages.

(See Unit-III, Page 119, Q.3)

- (b) Define HiperLAN technology.

**

- (c) What are the limitations of mobile TCP ?

(See Unit-IV, Page 193, Q.15)

- (d) Write and explain the four major challenges for implementing wireless LAN.

(See Unit-III, Page 118, Q.2)

Or

Draw and explain the architecture of 802.11a, 802.11b and 802.11g.

(See Unit-III, Page 141, Q.29)

Unit-IV

4. (a) Explain deregistration process in mobile IP.

(See Unit-III, Page 174, Q.53)

- (b) Briefly explain applications of mobile IP.

(See Unit-III, Page 174, Q.52)

- (c) Define mobile TCP with its objectives.

(See Unit-IV, Page 191, Q.10)

- (d) Write and explain seven major differences between MANET and wireless sensor networks.

**

Or

Draw a figure and explain the operation of mobile IP.

(See Unit-III, Page 163, Q.45)

Unit-V

5. (a) Write any five viruses names and purpose.
(b) What are worms ? How to prevent them ?
(c) Why Trojan horse is harmful for computer ?
(d) Explain intrusion detection system. How it works ?

**

**

**

**

Or

Explain firewall and what are the different types of firewalls.

(See Unit-I, Page 52, Q.55)

RGPV**IT-702**

B.E. (Seventh Semester)
EXAMINATION, Dec. 2016
WIRELESS AND MOBILE COMPUTING

- Note :** (i) Answer five questions. In each question part A, B, C is compulsory and D part has internal choice.

**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus

compulsory and D part has internal choice.

- (ii) All parts of each questions are to be attempted at one place.
- (iii) All questions carry equal marks, out of which part A and B (Max. 50 words) carry 2 marks, part C (Max. 100 words) carry 3 marks, part D (Max. 400 words) carry 7 marks.
- (iv) Except numericals, Derivation, Design and Drawing etc.

Unit-I

1. (a) What is antenna ? **
- (b) Write different types of fading. Explain how each one typically occurs. **
- (c) What is handoff ? Explain different handoff initialization techniques with suitable diagram. (See Unit-IV, Page 219, Q.35)
- (d) Discuss the various effects of mobile radio propagation. (See Unit-I, Page 215, Q.32)

Or

Compare between SDMA, TDMA, FDMA and DAMA. **

Unit-II

2. (a) What is UMTS ? **
- (b) What do you understand by mobility management ? **
- (c) What is DECT and TETRA ? **
- (d) Describe the GSM network architecture in detail with the help of diagram. (See Unit-IV, Page 227, Q.43)

Or

Discuss the different types of mobile service provided by GSM. Explain each in brief. (See Unit-IV, Page 225, Q.42)

Unit-III

3. (a) What is IEEE 802.11 ? (See Unit-III, Page 134, Q.19)
- (b) Write about MAC sub layer. (See Unit-III, Page 151, Q.36)
- (c) What will be the impact on piconet in Bluetooth devices are connected to mobile units ? Explain. **
- (d) Draw and explain HIPERLAN-2 protocol stack. **

Or

What are the MAC services of IEEE 802.11 that are not provided in the traditional LAN such as 802.3 ? (See Unit-III, Page 155, Q.38)

Unit-IV

4. (a) What is WAP ? **
- (b) Explain how mobile internet (mobile IP) protocol works. (See Unit-III, Page 166, Q.46)

**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus

- (c) Write a brief notes on snooping TCP and mobile TCP.

(See Unit-IV, Page 192, Q.11)

- (d) What is DHCP ? How does DHCP support dynamic address allocation ? **

Or

Explain how routing is performed on ad-hoc networks.

(See Unit-III, Page 174, Q.54)

Unit-V

5. (a) What do you understand by term intrusion detection ? **
- (b) Write the different types of viruses. **
- (c) Write a brief notes on biometrics and authentication system. **
- (d) Define the three classes of intruders and mention and intrusion techniques to protect from the intruders. **

Or

What is firewall ? Explain the design principle of firewall. **



RGPV

IT-702 (GS)
B.E. (Seventh Semester)
EXAMINATION, Dec. 2017
Grading System (GS)
WIRELESS AND MOBILE COMPUTING

Note : (i) Attempt any five questions.

(ii) All questions carry equal marks.

1. (a) What is antenna ? Give its types and their radiation pattern. ** 7
- (b) Describe wireless propagation modes and explain how a signal radiated along these modes. ** 7
2. (a) What is fading ? Describe flat and selective fading. ** 7
- (b) What are the benefits of spread spectrum technique ? Explain DSSS and FHSS techniques. (See Unit-III, Page 146, Q.35) 7
3. (a) Name the main elements of the GSM system architecture and describe their functions. (See Unit-IV, Page 227, Q.43) 7
- (b) Give reasons for a handoff in GSM and the problems associated with it. What types of handover can occur ? 7
- (See Unit-IV, Page 240, Q.56)
4. (a) What do you understand by Wireless LAN ? Write advantages and disadvantages of WLAN. (See Unit-III, Page 119, Q.4) 7

**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus

- (b) Explain architecture of Bluetooth networks. Discuss various applications of Bluetooth network. ** 7
5. (a) Explain the following – 7
- Hidden and exposed terminals problems
 - Near and far terminals problems.

(See Unit-III, Page 126, Q.9)

- (b) Discuss the requirements and architecture of HIPERLAN-1. ** 7
6. (a) What is Ad-hoc network? Write characteristics and applications of Ad-hoc networks. (See Unit-III, Page 130, Q.14) 7
- (b) Describe snooping TCP. Write advantages and disadvantages of snooping TCP. (See Unit-IV, Page 190, Q.8) 7
7. (a) Explain intruder. Discuss intrusion detection techniques. ** 7
- (b) Write and explain firewall design principles.

(See Unit-I, Page 47, Q.50)

8. Write short notes on any two – 14

- DHCP **
- CDMA **
- World Wide Web **
- Trojan horse defence. **

RGPV

IT-7003 (CBGS)
B.E. VII Semester
EXAMINATION, November 2018
Choice Based Grading System (CBGS)
WIRELESS AND MOBILE COMPUTING

Note : (i) Attempt any five questions.

(ii) All questions carry equal marks.

- What is antenna? Give its types and their radiation pattern. **
 - Describe wireless propagation modes and explain how a signal radiated along these modes. **
- What is fading? Describe slow and fast fading. Also explain large scale and small scale fading. **
 - Why a specialized MAC is used in wireless network? Explain the problems associated with 'hidden and exposed terminals' and 'near and far terminals'. (See Unit-III, Page 127, Q.10)
- What are the various mobile services provided by GSM? Explain each in brief. (See Unit-IV, Page 225, Q.42)
 - Give reasons for a handover in GSM and the problems associated with it. What types of handover can occur? (See Unit-IV, Page 240, Q.56)

**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus

- Explain GPRS architecture and reference model. **
 - Explain the protocol architecture of DECT system. **
- What do you understand by wireless LAN? Write down some important requirements for wireless LANs. (See Unit-III, Page 117, Q.1)
 - Explain various IEEE 802 standards and basic architecture of IEEE 802.11 standard. (See Unit-III, Page 141, Q.28)
- What is mobile IP? Explain the basic operation of mobile IP. (See Unit-III, Page 163, Q.45)
 - What is the basic purpose of DHCP? How can DHCP be used for mobility and support of mobile IP? Explain. **
- With advantages and disadvantages, explain indirect TCP. Explain snooping TCP. (See Unit-IV, Page 190, Q.7)
 - What is an intruder? Describe the classification of intruders. **
- Write short notes on any two of the following –
 - Firewall design principle (See Unit-I, Page 47, Q.50)
 - HIPERLAN **
 - WSN. **

RGPV

IT-7003 (CBGS)
B.E. VII Semester
EXAMINATION, November 2019
Choice Based Grading System (CBGS)
WIRELESS AND MOBILE COMPUTING

Note : (i) Attempt any five questions.

(ii) All questions carry equal marks.

- Attempt all the parts – $2 \times 7 = 14$
 - Describe the antenna, antenna types and its radiation pattern. **
 - Defined the role of multiple access technique in wireless network. Also explain SDMA, DAMA, PRMA. **
- Attempt all the parts – $2 \times 7 = 14$
 - With the help of diagram, explain system architecture of GSM. Also show the interfacing between different subsystems along components of each subsystem. (See Unit-IV, Page 230, Q.47)
 - Describe the process of call handover when a mobile station move. (See Unit-IV, Page 241, Q.58)
- Attempt all the parts – $2 \times 7 = 14$
 - Explain the IEEE 802.11 protocol architecture with a neat diagram. (See Unit-III, Page 142, Q.31)
 - Discuss the features of bluetooth enabled devices and their security features. **
- Attempt all the parts – $2 \times 7 = 14$

**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus

- (a) Differentiate between traditional TCP and mobile TCP. Also explain the working of indirect TCP. **(See Unit-IV, Page 192, Q.12)**
- (b) Illustrate the TCP connection timeout freezing of transmission protocol. **(See Unit-IV, Page 194, Q.17)**
5. Attempt all the parts – $2 \times 7 = 14$
- (a) What are the components of a virus ? Discuss the four phases in the life time of a virus. **
- (b) Explain the design principle of firewall and its types. **(See Unit-I, Page 52, Q.55)**
6. Attempt all the Parts – $2 \times 7 = 14$
- (a) Illustrate the problem of near and far terminals in medium access control. **(See Unit-III, Page 128, Q.11)**
- (b) Explain the protocol architecture of hiperlan and also mark the advantage and disadvantage of hiperlan. **
7. Attempt all the parts – $2 \times 7 = 14$
- (a) Compare and contrast I-ICP, snooping TCP and mobile TCP. **(See Unit-IV, Page 192, Q.13)**
- (b) Explain the following with respect to mobile IP –
- (i) IP packet delivery
 - (ii) Agent discovery
 - (iii) Registration.
- (See Unit-III, Page 174, Q.51)**
8. Attempt all the parts – $2 \times 7 = 14$
- (a) Draw the explain GPRS architecture. **
- (b) How authentication is possible in wireless LAN ? List and discuss the possible attacks on such network. **(See Unit-III, Page 120, Q.5)**



**Now, according to new revised syllabus of R.G.P.V., it is not included in syllabus